

Open Research Online

The Open University's repository of research publications and other research outputs

The Disposition Effect, Trading Biases, and Cognitive Reappraisal

Thesis

How to cite:

Grayson, Paul (2017). The Disposition Effect, Trading Biases, and Cognitive Reappraisal. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2017 The Author

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's [data policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

THE OPEN UNIVERSITY
Faculty of Business and Law
Department of People and Organisations

PAUL J. GRAYSON

MA (Oxon) – PPP (Psychology with Philosophy)

MSc – Evolutionary Psychology

PhD Thesis

**The Disposition Effect, Trading Biases,
and Cognitive Reappraisal**

**Submitted for the degree of
Doctor of Philosophy in Behavioural Finance**

30th November 2016

ABSTRACT

Millions of people across the globe invest in financial markets, hoping to increase their wealth or income. Investing is risky of course, but some investors seem to battle not only the market but also themselves. In recent decades behavioural economics has uncovered many patterns of trading behaviour which appear to work against the investors who display them, resulting in lower investment returns. For this reason, these patterns of behaviour are regarded as trading biases.

This thesis investigates two main themes that follow from the observation of trading biases in financial markets, both ultimately related to helping investors make better investment decisions. First, to what extent do individual investors reliably demonstrate trading biases? Second, why do these biases occur and can their impact be mitigated?

With respect to the first theme, is there evidence that some biases have trait-like characteristics, and predict investors' behaviour across a range of situations? This theme is related to the "state vs trait" argument from personality psychology. Applied here, the question is whether variation in trading behaviour is better explained due to environmental differences when decisions are being made (state), or because people differ in their tendencies to behave in certain ways (trait-like). If the latter, investors who tend to express biases would benefit from understanding their own decision-making patterns.

Note that the claim here is not that the biases are literally traits themselves, but that they have trait-like characteristics such as intra-individual stability and construct validity. The evidence presented in this thesis suggests that the biases examined possess these characteristics because their expression depends on some underlying behavioural tendencies; these underlying tendencies may be traits, though they are not directly investigated here.

With respect to the second theme, a suspected cause of many trading biases is the influence of irrelevant emotions on decision making. This thesis investigates evidence for the role of emotions in trading biases. It also investigates the use of cognitive reappraisal as a de-biasing technique (i.e. a method to reduce the level of bias displayed in trading decisions), which works by reducing emotions during trading.

The main bias which this thesis examines is the disposition effect. This occurs when investors are more eager to sell gains than losses, or stated another way, when they hold losses longer than gains. So, it is a bias that affects decisions about selling, based on the profit or loss made on each stock. Although the disposition effect is widely studied in behavioural finance, it has not been demonstrated to be a persistent pattern of an investor's trading behaviour. Experimental studies measure a mathematically defined disposition effect in the lab, but assume that this research can provide insight into the disposition effect observed in financial markets.

Meanwhile cognitive reappraisal, a form of emotion regulation, has been shown to reduce the disposition effect, but only in students and only using lab conditions that lack ecological validity. Furthermore, the mechanism for reappraisal's effect has not been investigated.

As well as investigating whether the disposition effect has trait-like characteristics, it also examines the “constituent biases” of the disposition effect, cutting gains and holding losses. Though the disposition effect is defined as a difference between selling gains and losses, the two are treated as two sides of the same behaviour rather than two separate trading patterns. This thesis argues that these constituent biases should be treated as separate biases, rather than two aspects of a unitary disposition effect. Building on the first theme, the constituent biases are also assessed for trait-like characteristics; building on the second theme, the effect of cognitive reappraisal on each constituent bias is examined.

This thesis contains three studies, two involving retail investors and one with an adult novice sample. All three involve measuring trading biases with an ecologically realistic trading simulation, played multiple times. The two studies with retail investors also include a disposition effect scale. Cognitive reappraisal is tested in one of the two studies with retail investors and in the study with novices.

I contribute to the literature in establishing that the disposition effect has trait-like characteristics aspects, by showing that it can be reliably measured in both an ecologically valid trading simulation and a self-report scale, and that these link to real-world trading behaviour. Furthermore, I show that the disposition effect is not a unitary bias but that its two constituent biases, cutting gains and holding losses, are independent of each other. They too have trait-like characteristics, and can also be reliably measured using the same ecologically valid trading simulation. However, levels of each bias in investors are not associated with each other. This contrasts with the conventional approach to the disposition effect that treats them as opposite sides of a unitary bias. I argue they are independent biases which are measured together in the disposition effect. Furthermore, as independent biases they are likely to be underpinned by different underlying traits.

I build on the existing literature by showing that the disposition effect occurs in retail investors using this trading simulation. I find that cognitive reappraisal reduces the disposition effect, while I make improvements in the external validity of this test in both the measurement of the disposition effect and participants used. I extend knowledge of emotion regulation and trading biases by showing that cognitive reappraisal reduces the disposition effect by decreasing the tendency to hold losses. Lastly, I show that cognitive reappraisal is not effective under the same conditions in novices (as opposed to retail investors). This reminds us of the merit of testing de-biasing techniques with greater ecological validity, and suggests that implementing de-biasing techniques in real world decision making may be more difficult than it first appears.

ACKNOWLEDGEMENTS

First and foremost, I want to thank my long-suffering wife Helen. I have been doing this PhD in one form or another for our entire marriage, indeed the entire time we have known each other! She has been patient beyond the limits any spouse should have to endure, while I have got it finished. I love you lots darling, and couldn't have got through it without you.

Thank you to my supervisors, Mark and Ben. Both have gone beyond the call of duty, for example accompanying me to Milan to recruit participants for my first study. Ben has continued to provide diligent advice even after he had left Open University, and I had moved to writing-up part-time. Mark's well-known expertise in the field has certainly opened doors and made parts of this thesis possible, and his calming advice and positivity often kept me going. Their detailed and insightful feedback has been invaluable in improving my studies, interpreting them, and presenting my work in this thesis.

Thank you to Jeff and Anna at Saxo Bank. They developed the Two-Index Game which this thesis uses, and setup the web hosting used for data collection from retail investors. In addition, they helped with many other ad hoc issues related to using or adapting the game.

Thank you to everyone else at the Open University whom I met along the way. In particular, thanks to Gareth Davies, Dan Richards, Ruslan Ramanau and Jo Locke, who all had a significant impact on my time at the Open University.

Finally, thanks to my parents for their support first when I decided to do a PhD, and then during some difficult times completing it.

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	xi
List of Figures	xv
1 Introduction.....	1
1.1 Introduction	1
1.2 Research questions and hypotheses	2
1.3 Methods, links between studies and research questions	6
1.4 Contributions.....	9
1.5 Structure of thesis	10
2 Literature review	13
2.1 The disposition effect	14
2.1.1 <i>Introduction to the disposition effect</i>	14
2.1.2 <i>Refuted explanations for the disposition effect</i>	20
2.1.3 <i>A prospect theory account of the disposition effect</i>	21
2.1.4 <i>The disposition effect as a stable trading behaviour</i>	27
2.1.5 <i>Underlying mechanisms of the prospect theory account</i>	30
2.1.6 <i>Questioning the role of prospect theory</i>	35
2.1.7 <i>Cutting gains and holding losses as stable trading behaviours and independent components of the disposition effect</i>	38
2.2 Emotions, emotion regulation, and decision making	39
2.2.1 <i>The impact of emotions on decision making</i>	40
2.2.2 <i>What is emotion regulation?</i>	48
2.2.3 <i>Conclusions about emotions, emotion regulation and decision making</i>	56
2.3 Approaches to Judgment and decision making research	58
2.3.1 <i>Two paradigms of decision making research</i>	59
2.3.2 <i>The implications of the two approaches for the role of emotions in financial decision making</i>	66
2.4 Improving external validity in experimental studies.....	68
2.4.1 <i>Methods of studying the disposition effect</i>	68
2.4.2 <i>Ecological validity of experimental studies</i>	70
2.4.3 <i>Participants in experimental research</i>	74

2.5	Conclusions and research questions	77
2.5.1	<i>Research motivations shared by all 4 research questions</i>	<i>77</i>
2.5.2	<i>First research question</i>	<i>81</i>
2.5.3	<i>Second research question</i>	<i>83</i>
2.5.4	<i>Third research question</i>	<i>85</i>
2.5.5	<i>Fourth research question</i>	<i>90</i>
3	Research Philosophy and Methods	92
3.1	Research philosophy.....	92
3.2	Research Design	97
3.3	Methods	101
3.3.1	<i>The Two-Index Game</i>	<i>102</i>
3.3.2	<i>Measurement of the disposition effect.....</i>	<i>107</i>
3.3.3	<i>Participants.....</i>	<i>111</i>
3.3.4	<i>Protocols.....</i>	<i>113</i>
3.3.5	<i>Participant Incentives</i>	<i>114</i>
3.4	Conclusion	116
4	Testing the nature of trading Biases using the two-index game	118
4.1	Intra-individual stability of trading Biases in the two-index game	119
4.1.1	<i>Intra-individual stability of the disposition effect</i>	<i>119</i>
4.1.2	<i>Intra-individual stability of cutting gains</i>	<i>123</i>
4.1.3	<i>Intra-individual stability of holding losses.....</i>	<i>125</i>
4.2	Convergent validity of the disposition effect	127
4.3	Discriminant validity across plays of the two-index game.....	128
4.3.1	<i>Testing discriminant validity with the two-index game</i>	<i>129</i>
4.3.2	<i>Analysis of correlation matrices from each study.....</i>	<i>131</i>
4.4	Discriminant validity controlling for trading frequency	135
4.4.1	<i>Analysis of correlation matrices from each study.....</i>	<i>136</i>
4.5	Conclusion	140
5	Analysis of the trading scale and its relation to scores from the two-index game.....	142
5.1	Description and preparation of scale data.....	143
5.2	Convergent validity of the disposition effect: Correlations between trading data and the game	146
5.3	Splitting the scale into factors	148
5.3.1	<i>Inter-item correlations and possible groups.....</i>	<i>148</i>
5.3.2	<i>Factor analysis method.....</i>	<i>151</i>

5.3.3	<i>Factor loadings and item retention</i>	152
5.3.4	<i>Eigenvalues and factor retention</i>	154
5.3.5	<i>Item reduction using factor analysis</i>	155
5.3.6	<i>10 items</i>	155
5.3.7	<i>9 items (item 7 dropped)</i>	158
5.3.8	<i>8 items (items 7 and 10 dropped)</i>	160
5.3.9	<i>7 items (items 6, 7 and 10 dropped)</i>	161
5.3.10	<i>6 items: 4, 5, 8, 1, 3, 9</i>	162
5.3.11	<i>Final factors produced</i>	164
5.4	Testing the independence of the factors.....	164
5.4.1	<i>6 items: 4, 5, 8, 1, 3, and 9, using direct oblimin rotation</i>	165
5.5	Correlations between the scale factors and the two-index game	167
5.5.1	<i>Correlations between factor scores and game scores</i>	169
5.5.2	<i>Correlations between scale scores and game scores</i>	171
5.5.3	<i>Convergent validity measuring cutting gains</i>	173
5.6	Conclusion	177
6	The effect of reappraisal on the disposition effect in retail investors. 179	
6.1	Method of the Milan study	180
6.1.1	<i>Experimental design & reappraisal intervention</i>	180
6.1.2	<i>Participant recruitment</i>	181
6.1.3	<i>Data analysis</i>	183
6.2	Overall disposition effect.....	185
6.3	The effect of reappraisal using plays 3 and 4	186
6.4	The effect of reappraisal using plays 2-4	190
6.4.1	<i>DE tested over plays 2-4</i>	190
6.4.2	<i>Marginal model of plays 2-4</i>	193
6.5	Conclusion	198
7	The effect of reappraisal on the disposition effect and emotions during trading in novices	199
7.1	Method of the OU study	201
7.1.1	<i>Experimental design</i>	201
7.1.2	<i>Reappraisal instruction</i>	204
7.1.3	<i>Questionnaires</i>	205
7.1.4	<i>Data management</i>	210
7.2	Overall disposition effect.....	212
7.3	The effect of cognitive reappraisal.....	214

7.3.1	<i>Manipulation checks</i>	214
7.3.2	<i>The effect of reappraisal on DE, PGR and PLR</i>	216
7.3.3	<i>The effect reappraisal on emotions</i>	217
7.4	Follow-up analysis to null results	220
7.4.1	<i>Effect of difficulty of reappraisal</i>	220
7.4.2	<i>Reappraisal targets used</i>	224
7.5	Conclusion	227
8	Summary of findings.....	229
8.1	Does the disposition effect have trait-like characteristics?	230
8.1.1	<i>The disposition effect will show intra-individual stability</i>	231
8.1.2	<i>The disposition effect will show convergent validity across multiple measures</i>	232
8.2	Do cutting gains and holding losses have trait-like characteristics?....	233
8.2.1	<i>Cutting gains will show intra-individual stability</i>	235
8.2.2	<i>Holding losses will show intra-individual stability</i>	236
8.2.3	<i>There will be discriminant validity between cutting gains and holding losses in the two-index</i>	236
8.2.4	<i>There will be discriminant validity between cutting gains and holding losses in the scale</i>	237
8.2.5	<i>There will be convergent validity for cutting gains between the two-index game and the scale</i>	238
8.2.6	<i>There will be convergent validity for holding losses between the two-index game and the scale</i>	239
8.2.7	<i>There will be discriminant validity between cutting gains and holding losses across the two-index game and the scale</i>	239
8.3	Does cognitive reappraisal affect the disposition effect and its constituent biases, when tested in retail investors under conditions of greater external validity?	240
8.3.1	<i>Investors will show a disposition effect in the two-index game</i>	241
8.3.2	<i>Cognitive reappraisal will reduce the disposition effect</i>	241
8.3.3	<i>Cognitive reappraisal will reduce holding losses but not affect cutting gains</i> 242	
8.4	Does cognitive reappraisal affect the disposition effect and its constituent biases, by changing emotions during trading, when tested in novices under conditions of greater external validity?.....	243
8.4.1	<i>4.1-4.3 The effect of reappraisal on the disposition effect, cutting gains and holding losses, in novices</i>	244
8.4.2	<i>4.4-4.5 The effect of reappraisal on emotions, and mediation of its effect on the disposition effect</i>	245

8.5	Summary of all four research questions	245
9	Contributions and Discussion	248
9.1	Contributions.....	249
9.1.1	<i>Does the disposition effect have trait-like characteristics?</i>	<i>249</i>
9.1.2	<i>Do cutting gains and holding losses have trait-like characteristics?.....</i>	<i>251</i>
9.1.3	<i>Does cognitive reappraisal affect the disposition effect and its constituent biases, when tested in retail investors under conditions of greater external validity? 253</i>	
9.1.4	<i>Does cognitive reappraisal affect the disposition effect and its constituent biases, by changing emotions during trading, when tested in novices under conditions of greater external validity?.....</i>	<i>255</i>
9.2	Implications and further research.....	257
9.2.1	<i>Trait-like characteristics of the disposition effect, cutting gains and holding losses 257</i>	
9.2.2	<i>The effect of cognitive reappraisal on the disposition effect, its constituent biases, and emotions during trading. Tested in experienced traders and novices under conditions of greater external validity</i>	<i>262</i>
9.3	Limitations and further research	271
9.3.1	<i>Trait-like characteristics of the disposition effect, cutting gains and holding losses 271</i>	
9.3.2	<i>The effect of cognitive reappraisal on the disposition effect, its constituent biases, and emotions during trading. Tested in experienced traders and novices under conditions of greater external validity</i>	<i>274</i>
9.4	Implications for practioners.....	280
9.5	Overall conclusions	282
	References	285
	Appendix 1: A mathematical explanation of the disposition using prospect theory.....	294
	Appendix 2: Risk aversion and loss aversion defined mathematically... 296	
	Appendix 3: Reappraisal instruction from the Milan study	297
	Appendix 4: Participant booklet from the OU Study (for reappraisal group participants)	298
	ID number:	298
	Outline.....	298
	Questionnaire 1	299
	Two-Index Game Tutorial and 1 st play.....	301
	<i>Entering your player name in the game</i>	<i>301</i>
	<i>Tutorial.....</i>	<i>301</i>

<i>1st play of the game</i>	302
<i>2nd play of the game</i>	302
Questionnaire 2	303
3rd play of the game & Questionnaire 3	306
<i>3rd Play of the game</i>	306
Questionnaire 3	307
4th play of the game & Questionnaire 4	310
<i>4th Play of the game</i>	310
Questionnaire 4	311
Debrief	315

LIST OF TABLES

Table 4.1 London study - correlations between DE on successive plays	120
Table 4.2 Milan study – correlations between DE on successive plays	120
Table 4.3 OU study – correlations between DE on successive plays.....	120
Table 4.4 Correlations of DE between play 1 with other plays in the Milan study	122
Table 4.5 Estimates of between-participant variance in DE	123
Table 4.6 London study – correlations between PGR on successive plays.....	124
Table 4.7 Milan study – correlations between PGR on successive plays.....	124
Table 4.8 OU study – correlations between PGR on successive plays.....	124
Table 4.9 Correlations of PGR between play 1 with other plays in the Milan study.....	125
Table 4.10 Estimates of between-participant variance in PGR	125
Table 4.11 London study – correlations between PLR on successive plays.....	126
Table 4.12 Milan study – correlations between PLR on successive plays	126
Table 4.13 OU study – correlations between PLR on successive plays.....	126
Table 4.14 Correlations of PLR between play 1 and other plays in the Milan study	127
Table 4.15 Variance in PLR scores accounted for by differences between participants	127
Table 4.16 London study - correlation coefficients between PGR and PLR in the two-index game.....	132
Table 4.17 London study - correlation p-values between PGR and PLR in the two-index game.....	132
Table 4.18 Milan study - correlation coefficients between PGR and PLR in the two-index game.....	133
Table 4.19 Milan study - correlation p-values between PGR and PLR in the two-index game.....	133
Table 4.20 OU study - correlation coefficients between PGR and PLR in the two-index game.....	134
Table 4.21 OU study - correlation p-values between PGR and PLR in the two-index game	134
Table 4.22 London study - correlation coefficients between PGR and PLR residuals in the two-index game of residuals after regression on trading frequency.....	137
Table 4.23 London study - correlation p-values between PGR and PLR residual in the two-index game after regression on trading frequency.....	137

Table 4.24 Milan study - correlation coefficients between PGR and PLR residuals in the two-index game of residuals after regression on trading frequency	138
Table 4.25 Milan study - correlation p-values between PGR and PLR residual in the two-index game after regression on trading frequency	138
Table 4.26 OU study - correlation coefficients between PGR and PLR residuals in the two-index game of residuals after regression on trading frequency	139
Table 4.27 OU study - correlation p-values between PGR and PLR residual in the two-index game after regression on trading frequency	139
Table 5.1 Items in the disposition effect scale	144
Table 5.2 Correlations between disposition effect scale scores and DE scores	147
Table 5.3 Disposition effect scale inter-item correlations	148
Table 5.4 Eigenvalues and factor retention for 10 items	155
Table 5.5 Communalities for 10 items	156
Table 5.6 Factor loadings for 10 items	156
Table 5.7 Eigenvalues and factor retention for 9 items	158
Table 5.8 Communalities for 9 items	159
Table 5.9 Factor loadings for 9 items	159
Table 5.10 Eigenvalues and factor retention for 8 items	160
Table 5.11 Communalities for 8 items	160
Table 5.12 Factor loadings for 8 items	161
Table 5.13 Eigenvalues and factor retention for 7 items	161
Table 5.14 Communalities for 7 items	161
Table 5.15 Factor loadings for 7 items	162
Table 5.16 Eigenvalues and factor retention for 6 items	162
Table 5.17 Communalities for 6 items	163
Table 5.18 Factor loadings for 6 items	163
Table 5.19 Eigenvalues and factor retention for 6 items with direct oblimin rotation ..	166
Table 5.20 Communalities for 6 items with direct oblimin rotation	166
Table 5.21 Factor loadings for 6 items with direct oblimin rotation	166
Table 5.22 2nd order factor loadings and communalities	167
Table 5.23 Factor loadings for 6 items with 2nd order factor	167

Table 5.24 Correlations between PGR scores and gain factor scores.....	170
Table 5.25 Correlations between PLR scores and loss factor scores.....	170
Table 5.26 Correlations between PGR scores and loss factor scores.....	170
Table 5.27 Correlations between PLR scores and gain factor scores.....	171
Table 5.28 Correlations between PGR scores and gain factor scale scores.....	172
Table 5.29 Correlations between PLR scores and loss factor scale scores.....	172
Table 5.30 Correlations between PGR scores and loss factor scale scores.....	172
Table 5.31 Correlations between PLR scores and gain factor scale scores.....	173
Table 5.32 Items from the scale relating to gains.....	173
Table 5.33 Correlations between PGR scores and all-gains scale scores.....	175
Table 5.34 Correlations between PGR scores and selling-gains scale scores.....	176
Table 6.1 Mean disposition effect overall and on separate plays.....	186
Table 6.2 Changes in trading biases by group.....	186
Table 6.3 T-test of mean change in DE by group.....	187
Table 6.4 T-test of mean change in PGR by group.....	189
Table 6.5 T-test of mean change in PLR by group.....	190
Table 6.6 Testing covariance structures for marginal model of DE scores.....	194
Table 6.7 Testing fixed effects in marginal model of DE scores.....	195
Table 6.8 Testing individual betas in marginal model of DE scores.....	196
Table 7.1 Positive and negative items in the PANAS.....	207
Table 7.2 Selected items from the PANAS-X used in the OU study.....	208
Table 7.3 Number of missing values per variable.....	211
Table 7.4 Disposition effect on each play.....	213
Table 7.5 Participants excluded by group.....	215
Table 7.6 Change in perceived responsibility by group.....	215
Table 7.7 T-test of change in perceived responsibility between groups.....	216
Table 7.8 Change in trading biases by group.....	217
Table 7.9 T-tests of change in trading biases between groups.....	217
Table 7.10 Change in positive and negative affect by group.....	218

Table 7.11 T-tests of change in positive and negative affect between groups	218
Table 7.12 Change in PANAS-X scales by group	219
Table 7.13 T-tests of change in PANAS-X scales between groups.....	219
Table 7.14 Change in other emotion items by group	220
Table 7.15 T-tests of change in other emotion items between groups	220
Table 7.16 Change in responsibility in the reappraisal group split by difficult rating ...	221
Table 7.17 Change in DE in the reappraisal group split by difficult rating	222
Table 7.18 Reappraisal targets and reappraisal difficulty rating	225
Table 8.1 Findings for trait-like characteristics of the disposition effect	231
Table 8.2 Findings for trait-like characteristics of cutting gains and holding losses	234
Table 8.3 Findings for the effect of cognitive reappraisal on the disposition effect and its constituent biases, tested in retail investors.	241
Table 8.4 Findings for the effect of cognitive reappraisal on the disposition effect and its constituent biases, and the role of emotions in reappraisal's effect, tested in novices.	244

LIST OF FIGURES

Figure 1.1 Links between studies, chapters, research questions and hypotheses in the thesis	8
Figure 2.1 Evaluating a gain and loss using the value function	24
Figure 2.2 Emotion-regulation strategies (from Koole, 2009)	49
Figure 3.1 Screenshot of the two-index game	102
Figure 5.1 Distribution of disposition effect scale scores (imputation 1).....	146
Figure 5.2 Venn diagram of main groups of inter-item correlations	149
Figure 5.3 Venn diagram of all significant inter-item correlations.....	150
Figure 6.1 Absolute DE scores over plays 2-4, split by group	190
Figure 7.1 Experimental design of the OU study	201
Figure 7.2 Summary of missing values.....	211
Figure 7.3 Disposition effect on each play.....	213
Figure 7.4 Change in responsibility by group.....	216
Figure 7.5 Change in responsibility in the reappraisal group split by difficulty rating ..	221
Figure 7.6 Change in DE split by difficulty rating	223
Figure 7.7 DE by group and reappraisal target on plays 3 and 4	226

1 INTRODUCTION

1.1 INTRODUCTION

In recent decades, there has been an explosion in research on decision making biases. Researchers in judgement and decision making have proposed many biases to explain behaviour and decisions which seem to be irrational or self-defeating. Some of the most economically important of these affect financial decisions, and one such bias is the disposition effect.

This bias has been recognized in academia since the seminal paper by Shefrin and Statman (1985). As you would expect, it is also recognized by investors in financial markets (for example Gross, p.150, 1982) although it is not usually known as the disposition effect by investors. This terminology originates from Shefrin and Statman who note it is a disposition where investors "sell winners and ride losers", meaning that they are eager to sell investments which have risen in value while being reluctant to sell investments which have fallen in value. The net effect is that investors are less likely to sell investments where they have made a loss than if they have made a gain.

Since Shefrin and Statman's paper, the disposition has been widely and robustly established in field studies, to the point where Barberis and Xiong (2009) note it is "one of the most robust facts about the trading of individual investors". Behavioural finance has established that the disposition effect occurs in financial markets across a huge range of countries and market types. Unfortunately, the bias does not usually help investors who display it. The disposition effect does not make sense on logical grounds: the fact that an investor has lost money on an investment does not mean it is a good investment for that investor to hold now. Empirical evidence backs up this reasoning, for example Odean (1998) finds that the winners investors sell outperform the losers they hold.

These studies have provided researchers with an observed phenomenon to explain. Some researchers have used field studies to find demographic correlates or investor characteristics which are associated with higher or lower levels of the bias (Feng and Seasholes, 2005, Dhar and Zhu, 2006). Others have attempted to explain the underlying psychology that drives the effect, often using experimental studies (Camerer and Weber, 1998, Summers and Duxbury, 2012). This thesis addresses two themes, which result from combining aspects of field studies and experimental studies.

1.2 RESEARCH QUESTIONS AND HYPOTHESES

Many biases have been proposed by researchers, but an underexplored area is the extent to which these biases can be viewed as trait-like phenomena, rather than state phenomena. This is the first theme: are some individuals more prone to biases than others, and would they continue to display this bias in a different context? This thesis applies this question to the disposition effect, and its constituent biases, cutting gains and holding losses.

Although field studies have shown conclusively that a disposition effect does occur in financial markets, is this a consequence of the environments where it is measured, or is it because some investors are prone to that pattern of behaviour? Is it defensible to talk of an investor's "trading personality"? This question also has implications for the conclusions we can draw from carrying out experimental work. Would individuals displaying a disposition effect in experimental studies also display one in financial markets: can the former shed light on the latter?

A further qualification is necessary here, since the claim in this thesis is not that the biases are fundamental aspects of a person's personality, in the way that traits such as extraversion and neuroticism are viewed by psychologists. The behavioural scope of the disposition effect is too narrow for it to be a fundamental aspect of someone's personality. However, it is argued that these biases have some characteristics of traits,

and are manifested as stable behavioural tendencies in trading environments. This leads to the first research question and its associated hypotheses:

RQ 1 - Does the disposition effect have trait-like characteristics?

H 1.1 - The disposition effect will show intra-individual stability

H 1.2 - The disposition effect will show convergent validity across multiple measures

The same question is applied to the "two sides" of the disposition effect: cutting gains and holding losses. Do they also have trait-like characteristics, and can they also be seen as stable behavioural tendencies of investors, rather than simply a product of the particular environment decisions are made in?

Although the disposition is defined and often measured as a difference in decision making between gains and losses, it may make more sense to treat these "two sides" as independent biases? So, this is also examined in the second research question and its hypotheses, which are:

RQ 2 - Do cutting gains and holding losses have trait-like characteristics?

H 2.1 - Cutting gains will show intra-individual stability

H 2.2 - Holding losses will show intra-individual stability

H 2.3 - There will be discriminant validity between cutting gains and holding losses in a realistic trading simulation

H 2.4 - There will be discriminant validity between cutting gains and holding losses in the scale

H 2.5 - There will be convergent validity for cutting gains between a realistic trading simulation and the scale

H 2.6 - There will be convergent validity for holding losses between a realistic trading simulation and the scale

H 2.7 - There will be discriminant validity between cutting gains and holding losses, between a realistic trading simulation and scale

The second theme investigates the role of emotions in decision making biases, and testing psychological interventions to reduce biases with an emphasis on greater external validity in an experimental context. Again, this is applied to the disposition effect and its constituent biases.

Emotions have been shown to affect our decision making in many ways. In dual process theory, decision making is modelled as comprising two systems, system 1 and system 2 (Sloman, 2002), contrasting rapid and often automatic reasoning against deliberative and objective reasoning. Emotions affecting decision making are a possible example of system 1 in action: they provide a rapid and psychologically salient heuristic on which to base decisions (Schwarz, 1990).

The most common psychological explanation of the disposition effect is its derivation from prospect theory (Kahneman and Tversky, 1979). However, more recent research casts doubt on this, and suggests that emotions may play a prominent role in the disposition effect (Lee *et al.*, 2008, Summers and Duxbury, 2012) and other biases related to the disposition effect such as loss aversion (Sokol-Hessner *et al.*, 2009).

There is a large literature on interventions which can be used to mitigate biases, increasingly popularised in the last decade by books such as "Nudge" (Thaler and Sunstein, 2008) and prominent organisations like the UK's Behavioural Insights Team. Consistent with a role of emotions in decision making biases, one effective technique that has emerged is the use of emotion regulation. This is hypothesized to work by controlling the level of emotions experienced when making decisions, so reducing the bias caused by emotions. Among emotion regulation techniques, cognitive reappraisal has been especially effective in improving decisions (Wallace *et al.*, 2009, Sokol-Hessner *et al.*, 2009, Lee *et al.*, 2010). This involves reappraising a situation to change its significance or meaning, such that the new appraisal no longer gives rise to the unwanted emotion(s).

However, a drawback of many of studies testing emotion regulation is that the experimental demonstration of cognitive reappraisal takes place within an artificial lab setup. While this is practically helpful and aids experimental control, it may impair the external validity of findings. This is the case with Lee *et al.* (2008) who demonstrate a reduction in the disposition effect using cognitive reappraisal: the experimental task used to measure the disposition effect lacks ecological validity, and uses a student sample rather than participant with experience of trading.

This is achieved by increasing the ecological validity of the trading environment used experimentally, and using retail investors rather than students as participants. The third research question and its hypotheses follow from this:

RQ 3 - Does cognitive reappraisal affect the disposition effect and its constituent biases, when tested in experienced traders under conditions of greater external validity?

H 3.1 - Investors will show a disposition effect in a realistic trading simulation

H 3.2 - Cognitive reappraisal will reduce the disposition effect

H 3.3 - Cognitive reappraisal will reduce holding losses but not affect cutting gains

Continuing with this theme, the last research question aims to directly test the involvement of emotions in the disposition effect, and the effect of cognitive reappraisal on the disposition effect. This is also done using the same instrument to measure trading, so the greater ecological validity of the trading environment is maintained. The fourth research question and its hypotheses are:

RQ 4 - Does cognitive reappraisal affect the disposition effect and its constituent biases, by changing emotions during trading, when tested in novices under conditions of greater external validity?

H 4.1 - Novices will show a disposition effect in a realistic trading simulation

H 4.2 - Cognitive reappraisal will reduce the disposition effect

H 4.3 - Cognitive reappraisal will reduce holding losses but not affect cutting gains

H 4.4 - Cognitive reappraisal will reduce negative emotions experienced during trading

H 4.5 - Changes in emotions during trading will mediate the effect of reappraisal

1.3 METHODS, LINKS BETWEEN STUDIES AND RESEARCH QUESTIONS

I adopt a positivist philosophy and quantitative methodology, following many previous studies of decision making biases, emotion regulation and the disposition effect. This thesis contains three experimental studies, two involving retail investors (the Milan and London studies) and one with an adult novice sample (the OU study). All three involve measuring the disposition effect with an ecologically realistic trading simulation (the two-index game) which is played multiple times. The two studies using retail investors also include a disposition effect scale, which asked about their attitudes and behaviour in trading situations where a disposition effect may take effect, and there is also analysis of data linking the two-index to financial markets, supplied by Saxo bank who developed the game. Cognitive reappraisal is tested in one of the studies with retail investors, and in the study with novices.

The studies are combined in different ways to answer the research questions above. The first theme, comprising the first and second research questions, is about whether these biases have trait-like characteristics. This is tested by showing that the biases demonstrate intra-individual stability, convergent validity and discriminant validity, across a realistic trading simulation, the scale, and the trading data. To test the validity of cutting gains and holding losses, the scale is also split into two subscales.

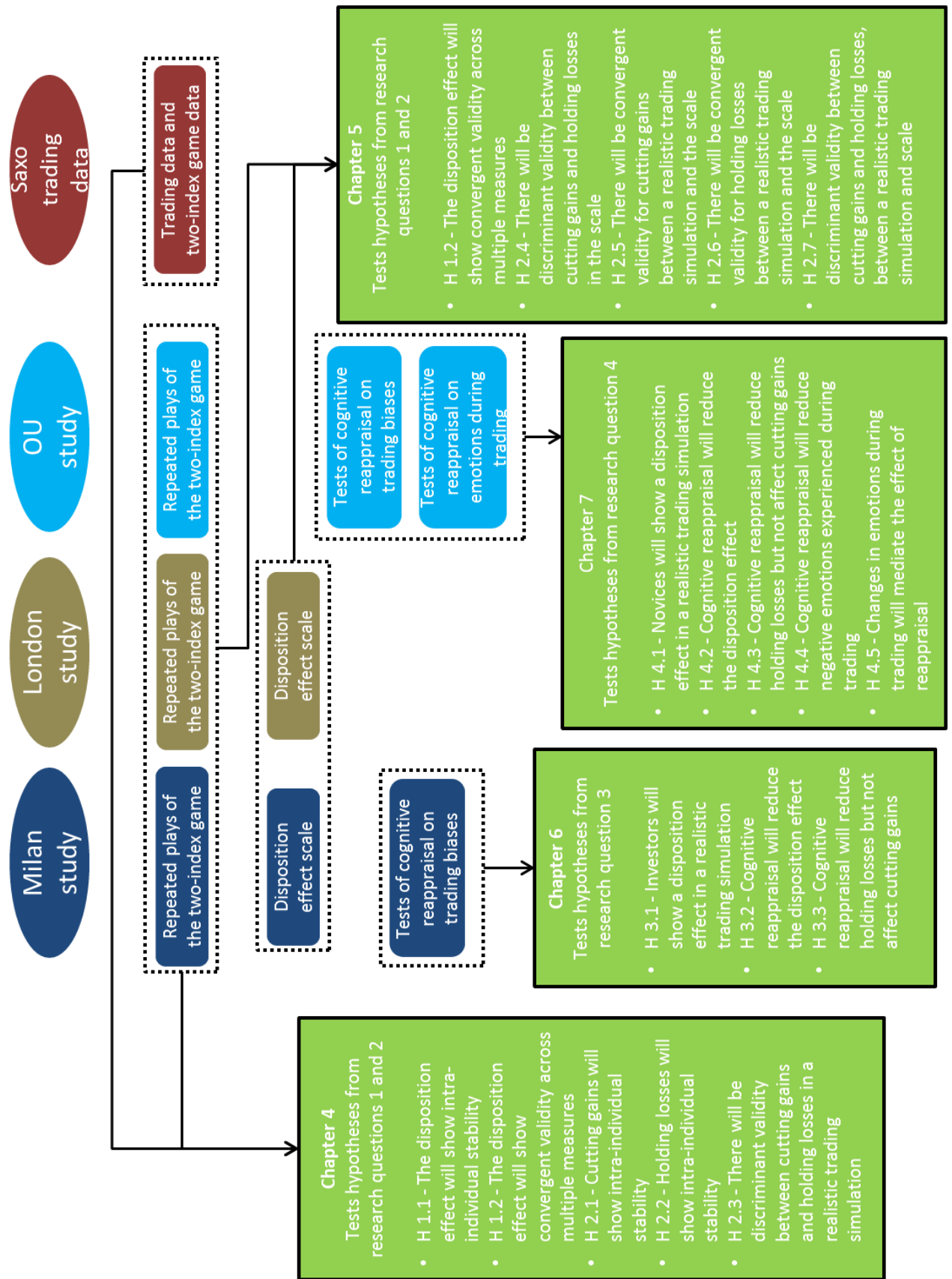
In the second theme, comprising the third and fourth research questions, cognitive reappraisal is tested experimentally by comparing a reappraisal group and a control group. The effect of emotions is tested in the OU study using emotion scales before and after reappraisal is used, and again compares reappraisal and control groups.

Figure 1.1 demonstrates how the studies are used to answer each research question, and in which chapter the material appears. Headings for each study (plus the Saxo data)

are at the top of the figure in ovals. Components of each study are shown vertically below each heading. The dashed boxes group together similar components across the studies. Finally, the arrows from each dashed group show which chapter they appear in, and which research question(s) and specific hypotheses they address in part or whole.

For clarity, the individual tests used to test each hypothesis are not shown. The primary purpose of this diagram is to show how the four studies (plus the Saxo data) relate to the hypotheses tested in this thesis, by showing what type of data was produced from each study and how different types of data are used to answer the four research questions. However, complete tables of tests for each research question are included in the summary of findings in chapter 8, and show which chapter contains those tests. A table is produced for each research question near the beginning of its respective section in chapter 8.

Figure 1.1 Links between studies, chapters, research questions and hypotheses in the thesis



1.4 CONTRIBUTIONS

This thesis makes a contribution with each of its research questions. It shows that the disposition effect has trait-like characteristics: it can be reliably measured in an ecologically valid trading simulation, the disposition effect from this simulation is consistent with one measured with a self-report scale, and it is also consistent with the disposition effect measured in real-world trading behaviour. This is strong evidence that the disposition effect is affected by a behavioural tendency in decision-making during trading which drives behaviour across a range of contexts. This implies it is not just a product of the situation.

A similar contribution is made with cutting gains and holding losses, showing that they too have trait-like characteristics. Both can be reliably measured in the lab using the same trading simulation. Holding losses also shows consistency across the trading simulation and a self-report scale, though cutting gains does not. The last part of this contribution is to show that only do they have trait-like characteristics, but that they are independent of one another. This entails that the disposition effect is not a unitary bias but that it is the combination of two constituent biases, cutting gains and holding losses, which should be regarded and researched as separate behaviours.

The third contribution shows that cognitive reappraisal can reduce the disposition effect, while testing this with increased external validity. This supports the work of Lee *et al.*, (2008), who demonstrated cognitive reappraisal's effect in a less ecologically valid environment. It also shows that cognitive reappraisal can be effective with people who have experience in financial markets. Finally, it includes the novel finding that cognitive reappraisal reduces the disposition effect by reducing holding losses, rather than cutting gains. This links back to the second contribution, where these constituent biases are shown to be independent behaviours. It also supports Richards (2012) who found that habitual reappraisal in investors may be associated with a lower disposition effect due to reduced holding of losses.

The fourth contribution is to raise questions about how psychological interventions may be tested with greater external validity, in contexts where expertise may be required to complete realistic tasks. There was no effect of cognitive reappraisal when tested with greater ecological validity, but with novices as participants rather than retail investors. Put another way, it may not be possible to use novices when attempting to study psychological biases with greater ecological validity, because the novices' lack of expertise makes it impossible for them to complete tasks properly in this environment.

1.5 STRUCTURE OF THESIS

This thesis is structured as follows. Chapter 2 reviews previous literature on the disposition effect, including proposed explanations for its occurrence such as an account based on prospect theory. It presents evidence that the disposition effect may be stable trading behaviour with trait-like characteristics. It goes on to critique the prospect theory account and suggest an explanation based on emotions during trading, which raises the question of whether cutting gains and holding losses are independent biases which also have with trait-like characteristics.

The next section reviews the wider links between emotions and decision making, and how emotion regulation has been used to reduce decision making biases. It briefly looks at different approaches to judgement and decision making research, before discussing external validity in previous experimental studies, and how this could be improved. It ends by reviewing how the literature examined motivates the research questions being answered in this thesis.

Chapter 3 begins by discussing the research philosophy which this thesis adopts. It proceeds to discuss the research design that is being used to investigate the research questions. For the first two research questions, this is the the multitrait-multimethod matrix (MTMM), developed by Campbell and Fiske (1959). This involves demonstrating intra-individual stability, convergent validity and discriminant validity, to establish the trait-like characteristics of the trading behaviours being explored. For the effect of

cognitive reappraisal in the third and fourth research questions, an experimental, repeated-measures comparison between reappraisal and control groups is used. This chapter concludes by reviewing some methods which are common across all three studies in the thesis. Paramount among these is an explanation of the realistic trading simulation (the two-index game), which is used to measure trading behaviour in the lab with greater ecological validity.

Chapter 4 begins the reporting of empirical results. Evidence for research questions 1 and 2 is split across chapter 4 and 5. Chapter 4 reports results of tests which use only the two-index game, or the two-index game with Saxo trading data. This chapter shows that all three biases are reliable, that the disposition effect shows convergent validity, and that cutting gains and holding losses have discriminant validity in the two-index game.

Chapter 5 presents evidence for research questions 1 and 2, where tests use the disposition effect scale in addition to the two-index game. First the scale is correlated with disposition effect scores from the two-index game to show convergent validity. Then to investigate the second research question, the scale is split into two subscales representing cutting gains and holding losses, and these subscales are shown to be independent of one another. Holding losses also shows convergent validity between the scale and the game, while both holding losses and cutting gains show discriminant validity between the scale and the game.

Chapter 4 and 5 combined show that the disposition effect, cutting gains and holding losses have trait-like characteristics. Furthermore, they show that the latter two should be treated as separate biases, independent of one another.

Chapter 6 addresses the third research question, and relates to the Milan study. It shows that the disposition effect can be reduced by cognitive reappraisal in retail investors, while also increasing the ecological validity with which it is tested. It also includes evidence that cognitive reappraisal does this by reducing the holding of losses.

Chapter 7 address the fourth research question and relates to the OU study. It is similar in design to the Milan study; however, it uses novices rather than retail investors, and measures emotions during trading to investigate the mechanism of cognitive reappraisal. It does not find an effect of cognitive reappraisal, and so poses some questions about when it is possible to test psychological effects with greater ecological validity.

Chapter 8 summarises the findings of this thesis. Each research question is dealt with in turn, and divided into the hypotheses which relate to it, and then finally the tests used to provide evidence for each hypothesis. The chapter draws conclusions for each research question.

Finally, chapter 9 discusses the contributions relating to each research question in more depth, and the implications for further research. Limitations of the work presented here are discussed, and then implications for practitioners, and finally overall conclusions for the thesis are drawn.

2 LITERATURE REVIEW

This literature review begins by reviewing the disposition effect. It discusses what the disposition effect is, why it matters, and evidence of its existence from field studies. Some common explanations for the disposition effect are examined, particularly one based on prospect theory, then the nature of the disposition effect is discussed and evidence is presented that suggests it may be stable trading behaviour which is expressed across time and contexts.

Research which questions the prospect theory account of the disposition effect is discussed, and this research leads to an alternative explanation of the disposition effect based on emotions during trading. It also suggests that cutting gains and holding losses are independent biases, and are underpinned by different emotions experienced during trading, rather than being two sides of a unitary disposition effect.

Building on the link to emotions during trading, the next section examines this area of research more widely. It reviews the relationship between emotions and decision making, introduces emotion regulation, and examines the effect of emotion regulation on decision making, focussing on how emotion regulation can be used to reduce decision-making biases.

Two opposing paradigms in judgment and decision making research are briefly reviewed: the heuristics and biases (HAB) approach, and naturalistic decision making (NDM). This locates the research in this thesis in the heuristics and biases paradigm; however, it also discusses the strengths of naturalistic decision making. A motivation of this thesis is to bridge the two paradigms, by testing hypotheses about the disposition effect experimentally like much heuristics and biases research, but with the greater external validity that research in naturalistic decision making often provides.

Next, two aspects of external validity in experimental studies are discussed: the ecological validity of the experimental setup, and the participants used. These aspects

are explored in relation to previous experimental studies on the disposition effect, and how they could be improved.

The chapter ends by drawing together on the literature reviewed in this chapter, and explaining how it motivates the research questions and hypotheses answered in this thesis. Later chapters expand on how these hypotheses are tested.

2.1 THE DISPOSITION EFFECT

2.1.1 Introduction to the disposition effect

The disposition effect is a decision-making bias, which is chiefly studied and observed in a financial context. A simple characterisation of the disposition effect is a trading pattern where investors “sell winners and ride losers” (Shefrin and Statman 1985), that is, they are eager to sell stocks which have gained in value, while being reluctant to sell stocks which have fallen in value. This has several corollaries which can be measured: losses will be held longer than gains (on average); the base rate probability of a stock being sold, given that one knows an investor has sold a stock of some sort, will be higher for a stock in a loss position than a flat or gain position. Since Shefrin and Statman’s paper, the disposition effect has been documented extensively in empirical investor studies (e.g. Odean 1998; Dhar and Zhu 2006), and also experimentally (e.g. Weber and Camerer 1998; Lee, Park *et al.* 2008).

Studies of the disposition effect use some financial terminology, which this thesis will also use. Buying a quantity of shares in a stock (i.e. shares of a company) is known as opening a position. Owning shares once they have been bought, but before they have been sold, is called holding a position. Selling those shares is called closing a position. Shares which are currently being held are an open position, while those that have been sold are a closed position. Shares which have risen in value since purchase are known as gains and those which have fallen in value are called losses. The money made or lost on a position is the payoff. Gains or losses which have not been sold yet are paper gains and losses, while those that have been sold are closed gains and losses.

During trading, there are two basic types of decisions to be made at any point: whether to purchase a stock, and following a purchase, whether to hold a position or close it. The disposition effect is a bias which affects the decision to hold or close positions. People who trade with a disposition effect tend to sell gains more quickly than losses, or conversely, they hold losses for longer than gains. This is also known as “cutting gains and holding losses”. A typical finding is that the relative probability of gains being sold is between 50% and 100% higher compared to losses (e.g. Odean, 1998).

2.1.1.1 Why is the disposition effect interesting?

There are two main motivations for studying the disposition effect. The first is that it is interesting, simply as a psychological and economic phenomenon. Why should investors consistently show a bias in how they treat gains versus losses? Economic theory dictates that this is an irrational bias, which should not occur. The second motivation has more practical application: people who trade with a disposition effect have lower returns than people who don't (Odean, 1998), and this effect is usually greater in retail investors than professional investors. These two motivations are discussed in more detail below.

The disposition effect breaks some basic assumptions in economics, and can be viewed as a persistent deviation from normatively rational models of decision-making. Investors are assumed to try to maximise their wealth. The strong form of the efficient market hypothesis (Fama, 1970) predicts that past price changes do not predict future ones: the market does not consider what a stock was worth, but only what it is worth now.

If the market does not care about the cumulative price history of a stock, we can safely assume that the market does not consider whether an individual investor has made or lost money on a stock. Indeed, this perspective illuminates the irrationality of an investor allowing their previous purchase to affect their decisions about a stock. Investors with a disposition effect are asking “have I made money in the past?” when the question should be “what will make me money in the future?” Put simply, the disposition effect allows irrelevant information to influence trading decisions. Indeed, a loss position could be

seen as analogous to a sunk cost, and should have no input into decisions about holding or selling investments.

Taking the past purchase price into account when making selling decisions should at best have no effect on trading performance. However, irrelevant information such as the purchase price may crowd out useful information, leading to inferior decisions and lower returns. It is also possible that focussing some attention on a purchase price will hinder the use of any beneficial information or strategy which an investor does have. We can be agnostic about an investor's strategy: the key point is that whatever strategy an investor has, focussing on the 'sunk cost' of any money already lost should not be one of them.

As mentioned, the strong form of the efficient market hypothesis assumes there is no predictive information in previous stock movements. In fact, there is much evidence that the strong form of the efficient market hypothesis is not completely correct. For example, effects due to momentum, weather, seasonality and many others have been found. However, in the case of momentum effects, these still adversely affect the disposition effect investor. By selling gains too quickly, they tend to miss out when stocks they have sold continue to rise. By holding losses too long, they incur further losses as the price of the stock they failed to sell continues to fall. In support of this argument against the disposition effect, Odean (1998) finds that stocks which investors sell outperform the ones they hold, and estimates a difference in performance between them of 3.4% per year – a substantial difference in annual return. Supporting this, Seru *et al.* (2010) find that trading with a disposition effect results in poorer investment returns, whereas investors without a disposition effect do not incur this penalty.

The disposition effect is also recognised as problematic amongst financial researchers and professionals. For example Gross (p.150, 1982) quotes a stockbroker describing the behaviour of his clients:

"Many clients, however will not sell anything at a loss... (this) has probably wrought more destruction on investment portfolios than anything else"

Glick (1957) commented on the discussion of this problem by professional traders. They note how gains that are sold soon after becoming gains will only ever yield small profits, whereas a small unsold loss can grow to a very heavy loss:

"Small profits and large losses is an expression oft repeated by traders... (it) is control of losses which constitutes the essential problem"

Despite some awareness of the disposition effect, it persists and has been demonstrated enough times and in enough contexts to be considered a reliable feature of investor behaviour. Given the economic cost to those it afflicts, an explanation of its causes would be welcome, and once its causes are understood, interventions to mitigate its occurrence might be possible.

A disposition effect may also prevent an investor from taking advantage of other investment opportunities. Assuming they have limited capital, refusing to sell losses ties up capital which could otherwise be invested elsewhere. It was explained above that focussing on purchase prices could distract an investor's attention from useful information they may have about when to close a position. This is another way in which a focus on purchase prices can be detrimental to investors, by distracting them or preventing them from opening positions in other stocks.

Investors with a disposition effect may also lose by paying higher capital gains tax than necessary. This was first noted by Shefrin and Statman (1985). They reasoned that holding stock losses results in an increase in capital gains tax, since the unrealised losses cannot be used to offset gains on other stocks sold. Supporting this, Odean (1998) shows that in December (the last month of the U.S. tax year) the aggregate disposition effect reverses; he attributes this to strategic selling of losses to reduce capital gains tax. However, he still demonstrates a robust disposition effect overall. The conclusion is that immediately before the tax deadline, an imminent tax bill is enough to overcome whatever is driving the disposition effect, but not at other times during the year.

Another theoretical argument against trading with a disposition effect is an increase in the likelihood of going bankrupt, or at least ceasing to trade after losing all available trading capital. A disposition effect increases risk, due to prolonged holding of losses if they do not return to breakeven. These losses may grow ever larger, until they consume much or all available capital. Alternatively, a combination of intermittent large losses, resulting from holding each of them too long, may eventually consume all trading capital.

The disposition effect may also have a detrimental effect in examples of gains and losses outwith a trading environment, for example selling a house. An important example relevant to most of the general population is Genesove and Mayer (2001), who found a disposition effect exists in the housing market. A house which has fallen in value will not increase in price simply to allow the home-owner to avoid a loss: the logic here is the same as the efficient market hypothesis in financial markets. Therefore, delaying a house sale due to a capital loss is irrational in the same way as delaying a stock sale because of a capital loss. (Genesove and Mayer also controlled for home owners who could not sell at a loss due to negative equity, so they only compared home owners who were able to sell for a loss, but chose not to.) Home owners who refuse to accept market price and delay selling may needlessly prolong their move, incurring financial costs and detrimental practical effects on their lives.

This result suggests that financial markets may simply allow the disposition effect (which is a bias affecting how gains and losses are treated) to be expressed more often. In financial markets, it is common for assets to be bought with the intention of selling for profit later in the short or medium term, and the purchase price functions as a salient reference point from which to perceive gains and losses. However, whenever these characteristics apply to another situation, a disposition effect may also be expressed.

2.1.1.2 *Field studies of the disposition effect*

The disposition effect is a very robust phenomenon in financial markets. Indeed, Barberis and Xiong (2009) note it is “one of the most robust facts about the trading of individual investors”. Field studies of the disposition effect typically analyse very large datasets of trading records, acquired from trading brokers or directly provided by trading exchanges. This effect has been demonstrated in many types of financial markets, such as stocks, bonds, mutual funds and commodities. It is seen in all types of investors (retail, financial, government, foreign), and across many countries, and in countries across the world.

For example, Schlarbaum *et al.* (1978) analysed 2,500 retail investor accounts of mutual fund purchases and redemptions. They found that about 60% of redemptions were gains, and 40% were losses, meaning there were 50% more gains sold compared to losses. Their interpretation was that retail investors were skilled at picking stocks, resulting in more gains being available to close. However, Shefrin and Statman (1985) had a different interpretation. They hypothesised that retail investors had as many losing funds, but were less likely to redeem them, and they named this tendency the disposition effect, after the “disposition” to sell winners and hold losers. (The name has persisted in academia, despite being rather un-descriptive of the bias).

These studies drew attention to a possible effect in financial markets, but the effect was conclusively established by Odean (1998), who analysed some 10,000 accounts of share investors in the U.S. from 1987 to 1993. He analysed a very large number of trading accounts, over a long period of time and varying market conditions. While meticulously ruling out alternative rational explanations (discussed below), Odean still found that gains are around 50% more likely to be sold than losses.

The disposition effect has been demonstrated many times since, in a wide range of geographical locations, markets and types of investor, for example: Barber *et al.* (2007) with individual and institutional investors in Taiwan; Brown *et al.* (2006) with individual

and institutional investors trading IPOs in Australia; Kaustia (2010) with private investors in Finland; and Locke and Onayev (2005) with commodity future traders. Grinblatt and Keloharju (2001) analyse every stock market investor in Finland, and demonstrated the disposition effect not only in retail investors, but also professional investors such as financial and non-financial corporations, and in government bodies. Heisler (1994) finds the disposition effect in professional investors in US Treasury bonds. The disposition effect has also been demonstrated in emerging markets such as China (Feng and Seasholes, 2005), and Taiwan (Shu, 2005).

Although the effect is usually studied in financial markets, there is evidence that the underlying bias has more widespread effects. The disposition effect may affect decisions whenever it is possible to frame scenarios in terms of gains or losses. A related financial example is Heath *et al.* (1999), who found a disposition effect in exercise of stock options. Exercise of stock options doubled when the stock price rose above its highest point in the preceding year.

2.1.2 Refuted explanations for the disposition effect

Several explanations of the disposition effect are consistent with a rational investor displaying one, despite the arguments set out above. These include: mean reversion of stock prices, portfolio rebalancing, and trading costs. These explanations are expanded on briefly below, with evidence that they cannot account for the disposition effect in practice.

A belief in mean reversion entails that stocks will oscillate around a mean price. So, stocks which have recently gained in price are expected to fall, and stocks which have recently decreased in price are expected to rise. This implies that it is rational to sell gains before they can fall in value, and hold losses to benefit from their subsequent rise in value. However, Odean (1998) finds that analysing stock performance shows the opposite: stocks which investors sell outperform those they hold. This shows that, if these investors are making decisions based on a belief in mean reversion, their belief in

mean reversion is mistaken. The gains they sell would be more profitable to hold than the losses which they choose to hold.

It could still be the case that investors believe, falsely, that mean reversion is true and is a good basis for a trading strategy. However, other studies have shown that the disposition effect still occurs even when participants know mean reversion is false. In an experimental study, Weber and Camerer (1998) inform participants that each stock has fixed probability of price rise or fall, consistent over many periods: the disposition effect still occurs. Lee *et al.* (2008) explicitly give participants forecasts about price movements, but the disposition effect still occurs.

Portfolio rebalancing involves trading to maintain fixed percentages of capital in certain stocks or stock sectors. Stocks which have risen in price will take up a larger percentage of capital, while stocks that have fallen will take up a smaller percentage of capital. The action to remedy this is to sell some of the gains, while holding more losses. However, Odean (1998) controls for this investment strategy and still finds a disposition effect.

Trading costs are a stepped cost: they are not completely proportional to the amount traded, having some fixed elements. This means that trading costs can be higher as a percentage of smaller trades, and selling stocks which have fallen in value will tend to be smaller trades. An investor may be reluctant to incur these higher percentage trading costs, so avoid selling losses as a result. However, after controlling for this explanation, Odean (1998) still finds a disposition effect too.

2.1.3 A prospect theory account of the disposition effect

The predominant explanation now given for the disposition effect was proposed by Shefrin and Statman in their seminal 1985 paper, and derives the disposition effect from the framework of prospect theory (Kahneman and Tversky 1979). Note that this thesis is not seeking to support this explanation. It is presented as background knowledge to contrast with alternative hypotheses about the effect of emotions on the disposition

effect which will be introduced later. Evidence that calls into question a prospect theory explanation is discussed below in section 2.1.6.

Prospect theory proposes that decisions are made in two parts. The first part is called the framing phase (originally called the editing phase in Kahneman and Tversky's 1979 paper), where the decision is set into context. The second part is the evaluation phase where different courses of action are evaluated and the one with the highest expected utility is chosen.

The framing phase puts the possible outcomes into a form which simplifies evaluation and choice between them. In the disposition effect, framing occurs by mental accounting (elaborated by Thaler, 1985) which has two aspects. First, each stock is associated with its own mental account, which is "opened" at the time the stock is bought. This account is kept separately from other stocks and assets (i.e. non-aggregation of positions). Second, each account is judged relative to a reference price. This is initially the purchase price, and often remains at the purchase price until the stock is sold. For example, one stock is bought at £10; the mental account is zero and £10 is the reference price. If the stock falls to £9, the mental account of the stock will be a loss of £1. This is unaffected by any other investments held by the investor, and whether they have made gains or losses.

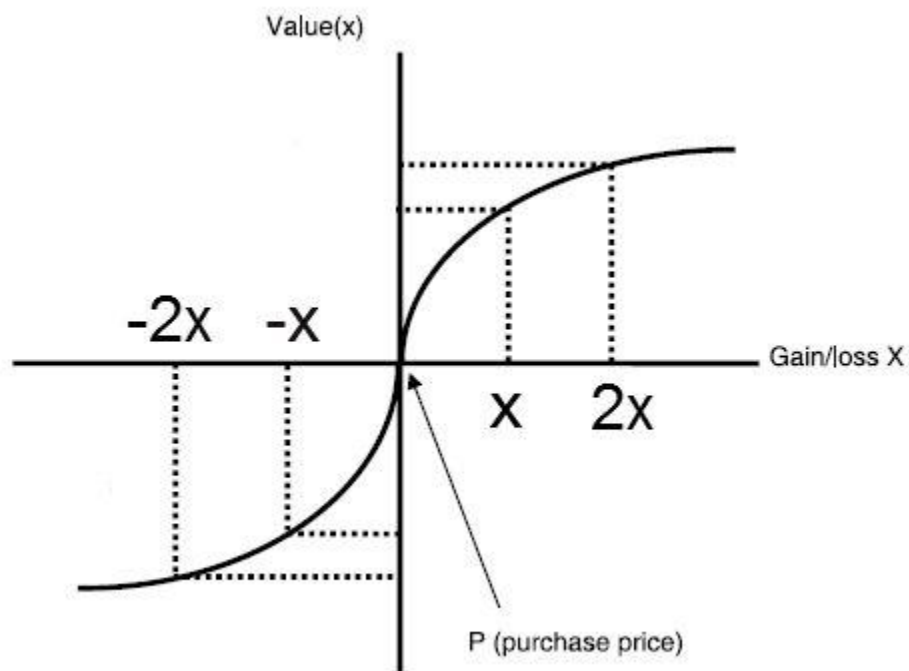
Thus, the balance on a stock's account is zero at the time of purchase and its future balance will reflect movements in the purchase price of the stock following purchase (assuming the purchase price remains as the reference price). As the price moves away from the purchase price by amount x or $-x$, the mental account is perceived as being in a gain or a loss position of magnitude x or $-x$. This has been implied in the introduction above without needing to be explicitly spelt out, since taking the purchase price as the reference price is an intuitive thing to do. Note also that this premise about mental accounting is not unique to prospect theory; however, what follows relies on mental accounting to take place, so mental accounting is a critical part of the explanation.

Once the position has been framed, courses of action can be evaluated. The decision the investor must make in this situation is whether to sell the stock immediately or to continue to hold it. However, instead of calculating expected value, prospect theory posits that expected value is transformed into expected utility, which is used as the basis for the decision. This transformation is called the value function and denoted " $v(x)$ ".

The expected utility from selling immediately is relatively simple to calculate, simply entering the gain or loss of the stock into the value function. This payoff will be achieved with 100% certainty if the stock is sold. Calculating the expected utility of holding is more complicated and requires an estimation of the probability distribution of possible future positions of the stock, and for the magnitude of those positions to be transformed by the value function to give an overall expected utility of holding which can be compared to selling. To simplify the example, in the standard explanation the only possible future outcomes are a future positive movement h or negative movement $-h$, both which have 0.5 probability of occurring. This mathematically embodies the assumption that future stock movements are independent of past movements.

Figure 2.1 shows how the value function is concave for gain positions and convex for loss positions. The x axis is the actual gain or loss of the stock since purchase, while the y axis represents expected utility the investor will experience from selling the stock. The shape of the value function means a change in position of the same magnitude on the x-axis has a different effect on expected utility depending on where the change is situated relative to the reference point. For simplicity and to exaggerate the effect, h here has been set to equal x . Thus, a positive movement of x from an existing gain of x results in a new position of $2x$. A negative movement of x returns the position to the origin (i.e. the purchase price). The reverse happens when starting from an existing loss position.

Figure 2.1 Evaluating a gain and loss using the value function



The disposition effect is produced when we translate these possible outcomes into expected utility, and the investor chooses the option which results in the greatest expected utility. When starting at a gain position of x , a further increase of x to $2x$ increases expected utility, as seen from reading off the y -axis in figure 2.1; however, the increase is not as large the decrease anticipated from returning to the break-even point. If these outcomes are equally likely, the expected utility of selling immediately will be higher than the average of the two future possible outcomes.

We have the reverse position for losses. A further movement of x back to the origin causes a larger increase in expected utility than the decrease in expected utility from an additional loss of x to an overall position of $2x$. So, if the future possible outcomes of holding are equally likely, expected utility will be increased by holding a loss rather than selling it, and selling a gain rather than holding it.

In practice, the model of binary outcomes of equal but opposite magnitude being known with certainty is obviously unrealistic. Investors may conceive of a probability

distribution of future movements, or they may just have an idea of what direction future price movements are likely to be, and estimate the probability of a further gain or loss from the current price. However, the prospect theory explanation is still applied despite these complications, as using the value function will still produce a higher expected utility for movements back to the origin compared to away from the origin, whether the position is originally a gain or a loss. The simplified version is usually presented as it is less complicated to explain. What matters however, is that there is a bias produced in favour of holding losses compared with gains, despite any complications, and that this is driven by the value function.

While investors will obviously not always estimate future outcomes of holding to be a gain or loss each with equal value and equal probability, what drives the disposition effect is that the value function creates a difference in behaviour towards gains and losses. Thus, they will *ceteris paribus* be expected to be more reluctant to sell losses, leading to longer holding of losses and lower base rate probability that a loss will be sold, and vice versa for gains. Please refer to appendix 1 for a more detailed mathematical explanation of how the disposition effect is derived.

2.1.3.1 Risk aversion and loss aversion terminology

There can sometimes be confusion about the terminology used to describe investors' behaviour. The tendency to sell gains in the disposition effect is sometimes referred to as risk aversion: investors forego potential further gains to avoid the risk of losing the existing gain (selling decreases risk since the payoff is known with certainty). The tendency to hold losses in the disposition effect is sometimes referred to as loss aversion: the investor takes on the risk of further losses for a chance to avoid having to take a loss at all (i.e. if the price increases back to the reference price, there will be no loss).

However, loss aversion can also describe another aspect of the value function, which is that losses give rise to greater absolute changes in expected utility than gains (Thaler,

1999). Most people refuse a bet of a 0.5 probability of a £110 gain and a 0.5 probability of a £100 loss, despite the expected value clearly being positive. However, the 'loss aversion' in the disposition effect is a different effect with a different cause. The cause of the risk (/ loss) aversion in the disposition effect is the greater expected utility of selling (/ holding) the stock compared to vice versa. Using loss aversion to describe this behaviour in the loss domain conflates two concepts: the behaviour produced by the decision maker comparing expected utility of small price changes from an existing position, and the absolute greater weighting of losses compared to gains.

Instead, these might be better referred to as gain-framed risk aversion and loss-framed risk propensity, or simply risk aversion and risk propensity. These are variations in an investors' risk appetite that are dependent on the value function, in conjunction with the framing of a stock's position. Risk aversion and risk propensity are both caused by greater **changes** in expected utility being generated by movements **back** to the reference price than away from it, whether the original displacement away from the reference price was negative or positive.

For gains, movement back to the reference price creates greater negative utility than the positive utility of movement away from the reference price, so the motivation is to prevent this by selling the gain. This decreases risk because the outcome of selling is certain. For losses, movement back to the reference price creates greater positive utility than the negative utility of movement from it, so the motivation of the decision maker is to encourage this by holding the loss. This increases risk because the certain outcome from selling is rejected. However, in neither case is the object to increase or decrease risk per se: the disposition effect as explained by prospect theory is driven by which option has the greatest expected utility. The effect on risk exposure is a consequence, not a cause, of this.

Loss aversion is better reserved to describe the greater expected utility of losses versus gains which have equal objective value. This type of loss aversion is **not** a part of the

standard explanation of the disposition effect by prospect theory which is set out above. Please refer to appendix 2 for a more detailed mathematical explanation of this distinction.

2.1.4 The disposition effect as a stable trading behaviour

The sections above have introduced the disposition effect. They have explained its significance, 'rational' explanations for the effect which have been refuted by previous research, and detailed the most common contemporary explanation based on prospect theory. Section 2.1.5 continues this by looking at the prospect theory account in more detail and section 2.1.6 discusses some of its weaknesses. However, now that enough background knowledge of the disposition effect has been covered, this section is an appropriate point to discuss the psychological nature of the disposition effect.

An interesting research question in behavioural economics is whether biases can be viewed as trait-like phenomena, rather than state phenomena, and this question can be applied to the disposition effect too. Is the disposition effect a stable behaviour which varies between individual investors? Put another way, does it make sense to refer to an individual as having a 'high disposition effect' or 'low disposition effect': if someone is found to express a high level of disposition effect in one context, are they likely to express a high level in a different context?

The account of the disposition effect given above suggests that it could be a stable behaviour. Prospect theory is a model of decision-making which can be applied to any context, and it explains the disposition effect being driven by a difference in how gains and losses are treated. Therefore, it should apply to any context where selling decisions are made about gains and losses. The difference between selling gains and losses is explained by the shape of the value function, which is modelled by only two parameters - α and β (refer to appendix 1 for more discussion of these). If these parameters vary between individuals but are stable within individuals, we would expect to see the disposition effect emerge as a stable behaviour across time and contexts.

Although most studies on the disposition effect involve stocks in financial markets, as discussed above, a few have shown its presence in other decisions. Heath *et al.* (1999) show it applies to stock options (admittedly not massively different from the trading of actual stocks). More persuasive is Genesove and Mayer's (2001) demonstration that people display the disposition effect when selling their house. Also, as discussed, the position of holding a loss is very like a sunk cost, where a loss already incurred should be disregarded as irrelevant to the decision being made. So, these examples suggest that the disposition effect could be a stable feature of an individual's decision making, which is merely most recognisable in trading stocks.

There is very extensive field research supporting the disposition effect in a wide range of financial markets and market participants. However, the nature of most of these studies is that they look at the behaviour of investors in only one market. This does not establish that the same investors would display a similar level of disposition effect in different situations. Meanwhile, Dhar and Zhu (2006) find significant variation in the disposition effect expressed in their population of investors. This is important for claiming the disposition effect to have a trait-like character; however, they do not establish that this variation is stable across contexts.

Some field studies though, do raise the possibility that the disposition effect is a stable feature of decision making in other ways. For example, Seru *et al.* (2010) show that over a 9-year period, the disposition effect of individual investors who are still actively trading declines. *Prima facie* this seems to be evidence that investors learn to avoid the disposition effect through their trading experience. However, the contribution of this type of learning to the decline is low. Most of the decline in disposition effect is in fact due to low ability (i.e. high disposition effect) investors ceasing to trade, which leaves predominantly investors with low disposition effects remaining in the population. They conclude that most of the 'learning' which takes place involves investors learning about their own trading ability. The implications for this thesis are that it provides evidence for the stability of the disposition effect in investors over time, despite extensive opportunities

for investors to reduce their disposition effect through trading. This comes with the caveats that the disposition effect is only measured with one method and with one sample, so state (as opposed to trait) influences on the disposition effect could also persist over time.

Other field studies suggest a stable disposition effect by making a link to other demographic or psychological variables which are already known to be stable. If the disposition effect is reliably correlated with some other stable variable, it suggests the disposition effect itself is stable. For example, Kadous *et al.* (2014) show that the disposition effect is correlated with two separate types of self-esteem. In addition, Richards (2012) shows that the disposition effect is correlated with cognitive style, as measured by the rational-experiential inventory (Norris and Epstein, 2009).

Experimental work has also found associations between demographic and psychological variables. The locus of control is a stable psychological variable which relates to how much control an individual believes they have on external events. Chiu (2001) finds that an external locus partly explains the disposition effect observed in their experiment.

Note that this thesis is not attempting to show that the disposition effect is a psychological trait. Kassin (2003) defines traits as 'habitual patterns of behaviour, thought, and emotion', which does imply that the disposition effect would qualify. However, in practice, the category of trait is reserved for more fundamental aspects of our psychological makeup. For example, Costa and McCrae (1992) use only 5 traits to define their Big Five model of human personality. Set against this, a difference between how gains and losses are sold in trading is not basic aspect of human psychology.

However, what is explored is whether the disposition effect does display some qualities that traits have, namely: is it stable within individuals while varying between individuals, and is it stable across different situation and contexts. These are referred to as 'trait-like characteristics'. If it does have these qualities, future research can investigate why: is it possible that more fundamental psychological variables which *are* traits underpin the

stable behaviour seen the disposition effect. However, even without this, answering the question of whether the disposition effect has trait-like characteristic will improve our understanding of how the disposition effect affects investors' trading behaviour.

2.1.5 Underlying mechanisms of the prospect theory account

Section 2.1.3 presented an explanation of the disposition effect derived from prospect theory. Prospect theory (Kahneman and Tversky, 1979, and in many subsequent papers) holds that rather than evaluate the expected monetary value of payoffs when making decisions, payoffs are converted into expected utility. This utility represents the subjective value of a payoff to the decision-maker. In addition, in prospect theory the conversion of expected value into expected utility is done in two stages. Firstly, a payoff is converted to a relative value by comparison with a reference point, and secondly, that relative value is transformed into utility using the value function. The value function shows diminishing sensitivity to movements away from the reference point for both gains and losses. In other words, the further a payoff is from the reference point, the smaller the subjective effect (i.e. utility) of further marginal changes in the monetary value of a payoff.

An explanation of the disposition effect based on prospect theory was first proposed by Shefrin and Statman (1985), and is commonly cited by other researchers studying the disposition effect. For example, Weber and Camerer (1998) begin their paper by demonstrating how the disposition effect could be produced using figures of the value function, and showing how payoffs from trading decisions would be evaluated with it. (This is similar to figure 2.1 in section 2.1.3). The effect of the value function transforming payoffs for gains and losses into expected utility is that, when future positive or negative movements are equally likely, the expected utility of selling a gain will be greater than holding a gain, and the expected utility of holding a loss will be greater than selling a loss. This combination results in a preference for selling gains and holding losses: in other words, it results in a disposition effect.

However, the account of the disposition effect from prospect theory is based on several underlying mechanisms, which will be discussed in further detail now. I also discuss experiments which disrupt these underlying mechanisms, so which are in turn expected to also disrupt the disposition effect.

2.1.5.1 Calculation of expected utility

A consequence of the difference between expected value and expected utility is that expected value treats all future price movements of equal size and equal likelihood as equally valuable, whereas the expected utility of a price movement depends on its relationship to the reference point. For example, as explained above, when an investor considers selling a gain, future further gains have less expected utility than future price falls, even if they are of equal size.

Although the value function transforms expected value into expected utility, decisions should still be sensitive to the likelihood and magnitude of future price movements. For example, in a gain position the magnitude and / or likelihood of a further gain would have to be somewhat greater than for a future loss to overcome the built-in bias in favour of selling gains. However, basing decisions on expected utility does not mean that the magnitude and likelihood of future price movements is ignored. (For a loss position, the likelihood and / or magnitude of a further loss would have to be greater than for a future gain, to overcome the built-in bias to hold losses).

Supporting this, Lee *et al.* (2008) found that negative expectations of future stock performance led to a greater probability of selling a loss. They manipulated participants' expectations of the future performance of a stock by using recent trend information and 'expert recommendations'. When 'experts' thought a loss would fall further in price, participants were more likely to sell the loss, as predicted by expected utility calculations. In addition, participants' beliefs about future performance mediated this effect. However, this still did not eliminate the disposition effect. So, this shows that decisions are sensitive to the changing likelihood of gains and losses, like calculations of

expected utility would be. However, beliefs about future performance cannot explain the disposition effect, since it still occurs when these beliefs change.

However, Lee *et al.* (2008) found that several experimental interventions did eliminate the disposition effect. One experimental group was given an expected value exercise before playing the investment game. This required participants to calculate the expected monetary return of various gambles, i.e. considering value (as in money) rather than utility, bypassing the putative value function and its transformation of value into utility.

The disposition effect was eliminated in the experimental group as predicted, and this elimination was independent of stock trend and forecast of future performance.

Manipulation checks implied a straightening of the hypothesised value function, so participants were making judgements corresponding more closely to choosing maximum expected value, rather than maximum expected utility. This suggests that it is the value function's transformation of expected value into expected utility that drives the disposition effect, which is consistent with the prospect theory explanation.

2.1.5.2 Mental accounting

The prospect theory explanation rests on the differential treatment of gains and losses, and mental accounting is necessary for the perception of gains and losses. This implies that mental accounting is necessary for the disposition effect to occur. Mental accounting in the disposition effect depends on using the purchase price as the reference price, and the non-aggregation of a stock position with other stock positions in an investment portfolio (and non-aggregation with an investor's net worth).

Sokol-Hessner *et al.* (2009) found that instructing participants to treat decisions as one of a series of trades that a professional trader may take, rather than treating them individually, significantly decreased loss aversion. This experimental manipulation may have had its effect by disrupting the non-aggregation requirement of mental accounting. While the dependent measure was loss aversion rather than loss-framed risk appetite

(i.e. holding losses), this also relies on mental accounting for framing to 'create' the losses that loss aversion is directed towards.

Kaustia (2010) found there is an abrupt jump in the likelihood of a stock being sold just at the point the stock breaks-even or becomes a small gain. This jump may appear to be consistent with the prospect theory account of the disposition effect, since it predicts an asymmetry in how gains and losses are treated. However, invoking the value function is not the simplest explanation. Mental accounting and qualitative differences in reactions to gains and losses could account for Kaustia's result without invoking the value function. For example, if people are more averse to selling at a loss, and sell at break-even or a small gain once they are given the chance, the jump found by Kaustia could also be predicted.

2.1.5.3 Reference price adaptation

Disrupting the use of the reference price should eliminate the disposition effect.

Adaptation describes the hypothetical movement of the reference price to take into account changes in the market price, either partly or fully. The prospect theory account above assumes that there is no adaptation so gain and losses are always calculated by referring to the purchase price.

Adaptation should change the choices made by changing the expected utility of the options. However, the effects of partial adaptation (i.e. the reference price only moves part of the way to the current price) are more complicated than simply eliminating the disposition effect. Incomplete adaptation will preserve the position of a stock as a gain or loss, but will have the effect of shifting the curve of the value function towards the current price, effectively steepening whatever gradient the curve had before. As the difference in expected utility between holding and selling is greater the greater the gradient of the curve, this would result in an increased propensity to hold losses / sell gains, and an increase in the disposition effect. Thus, incomplete adaptation is predicted to increase the disposition effect, whereas complete adaptation should eliminate it.

Lee *et al.* (2010) find that adaptation to losses is correlated positively with agreeableness and intellect, and negatively with conscientiousness (there was no relationship between adaptation to gains and any personality traits). Their work suggests that propensity to adaptation, of losses at least, is a stable personality feature. Personality traits thus could have a relationship with the disposition effect, mediated by the extent of adaptation. The correlation between personality traits and adaptation to losses but not adaptation to gains also reinforces the idea that reactions to gains and losses are categorically different. Of course, a difference in how gains and losses are treated is part of the disposition effect by definition.

2.1.5.4 The distinction between realised and unrealised positions

The prospect theory explanation assumes that utility is not experienced by an investor until a stock position is closed. That is, unrealised losses and gains produce utility when they are realised, but until this point they are only considered as a source of expected utility in the future. Removing the distinction between realised and unrealised positions should negate the disposition effect.

Barberis and Xiong (2009) developed a model showing that when there is no division between the utility gained from a realised loss and an unrealised one, prospect theory cannot account for the disposition effect except in some specific limited circumstances. The purpose of their paper was to establish whether the mental accounting assumption is necessary. Since it appears to be, future models need to include this premise and it needs to be verified experimentally.

A reservation about their model is that they assume stocks must have high expected returns when they are bought. They state that, given the inflection point of the value function (which produces loss aversion), investors should not buy a stock if this is not the case since they are averse to 50/50 "fair" gambles. However, what is actually necessary is that investors believe there are high expected returns (this doesn't have to be true, only that they believe it). Since investors have a large selection of stock to

choose from when buying, it seems possible that many or most stock purchases can be explained by random variation of subjective valuations away from the true value. Where investors overvalue a stock enough they will want to buy it. It is not clear if the results of their model would change by relaxing this assumption about high returns.

2.1.6 Questioning the role of prospect theory

The evidence above about the underlying mechanisms of a prospect theory account of the disposition effect are broadly supportive of this account. However, more recent work has undermined the prospect theory explanation by producing results that are inconsistent with it. In particular, they question the role of the value function in driving the effect by showing that selling decisions respond to a distinction between gains and losses, but do not respond to changes in their magnitude in the way prospect theory predicts. An alternative explanation proposed is that there is a qualitative difference in emotional reactions to gains and losses, which drives the disposition effect.

Kaustia (2010) shows that the probability of closing a stock position rises rapidly around the breakeven point, where a loss changes to a gain. The probability of selling is very insensitive to how large the loss is: what matters is whether it is a loss or it is not a loss. Similar results were found for gains, which are insensitive to the size of a gain over a large range of profits.

Prospect theory predicts a more gradual response as the position changes from a loss to a gain, which matches the S shaped curve of the value function. However, the conclusion drawn by Kaustia is that there is a sharp distinction between losses and gains, rather than a gradual one. The distinction between gains and losses does not appear to arise from the transformation of gains and losses by the value function, but from a qualitative difference between gains and losses.

Lehenkari (2012), Summers and Duxbury (2012) and Lee *et al.* (2008) have also undermined an explanation based on prospect theory, by finding results that should not

occur if prospect theory can fully account for the disposition effect. Prospect theory implies that any potential gain or loss should be assessed in the same way, regardless of how an investor came to hold that position. However, Lehenkari found that investors who inherited positions had smaller disposition effects than those who had opened the positions themselves. A tentative conclusion is that responsibility for having a gain or loss position is necessary to induce a disposition effect. This is inconsistent with prospect theory: there is no psychological input in prospect theory which could represent why an investor is presented with the trading options they have, or in what context.

Summers and Duxbury (2012) build on this work in an experimental study, and make a link to the involvement of emotions in trading decisions. They show that participants who are not responsible for holding a stock do not show a disposition effect when selling that stock. They conclude that simply having a gain or loss is not sufficient to produce the disposition effect, but that investors need to feel responsible for holding that stock in the first place. Again, this is inconsistent with prospect theory, which should apply to all gains and losses regardless of context.

Furthermore, Summers and Duxbury show that responsibility for trading decisions is associated with the experience of specific emotions, and conclude that these emotions are also necessary for the disposition effect to occur. Participants who were responsible for losses felt more regret than those who incurred losses but were not responsible for them. (Disappointment was also higher when incurring losses but did not vary with responsibility). For gains, experiencing elation was associated with more selling gains; however, the experience of elation was not linked to being responsible for a gain.

Summers and Duxbury conclude that an increase in regret, in response to being responsible for a loss, is what drives the disposition effect. In addition, experiencing elation is necessary to drive the selling of gains. Their work echoes another explanation proposed by Shefrin and Statman (1985), that the disposition effect is associated with investors seeking pride from selling gains, and avoiding regret by not selling losses.

Summers and Duxbury extend this by linking the experience of regret to responsibility for the loss incurred. Building on the work of Summers and Duxbury, Rau (2015) finds that two-person teams exhibit a higher disposition effect than individuals and rarely sell losses. Specifically, teams reporting high levels of regret leads to this behaviour.

In another experimental study which prospect theory cannot account for, Lee *et al.* (2008) found that instructing participants to use a form of emotion regulation to reduce the disposition effect. Participants were asked to assume the role of a stock broker, making selling/holding decisions on behalf of customers. This almost completely removed the disposition effect compared with a control group. Prospect theory cannot account for this as trading for another person should make no difference to how prospects are evaluated when compared to the reference price.

The type of emotion regulation used by Lee *et al.*, cognitive reappraisal, is defined by Koole (2009) as 'changing the subjective evaluations during emotionally significant events'. This often occurs by reframing a situation by changing its context. If emotions are involved in producing the disposition effect, improved emotion regulation by investors would be expected to reduce the effect, by controlling or reducing the influence of those emotions. So, this could explain the results of Lee *et al.*: trading on behalf of someone else may reduce the emotions experienced in trading. Richards (2012) has also found tentative evidence that cognitive reappraisal affects the disposition effect by reducing holding losses.

The results above suggest two things. First, prospect theory cannot account for these manipulations of the disposition effect and therefore it cannot produce a full account of how the effect is produced. Second, an account of the disposition effect is likely to include an explanation of how emotions during trading are involved, for example that a difference between the emotional response to gains and losses drives the difference in trading behaviour towards gains and losses.

2.1.7 Cutting gains and holding losses as stable trading behaviours and independent components of the disposition effect

The possibility that the disposition effect is driven by different emotional responses to gains and losses allows further exploration of the nature of the disposition effect, raised earlier in section 2.1.4.

The disposition effect has traditionally been viewed as one bias, identified by measuring a difference in the frequency that an investor sells gains compared with losses. Prospect theory predicts that it is a unitary bias, because the behaviour towards gains and losses are both driven by the curve of the value function. The more curved the value function, the higher disposition effect someone will express; however, the value function will be curved in both gain and loss domains, so both cutting gains and holding losses should increase.

In contrast, recent research suggests that cutting gains and holding losses could be two independent biases with distinct causes, which in combination can produce a disposition effect. In other words, two investors could have a similar disposition effect overall, but the disposition effect in one investor could be due to greater cutting gains than average, while in another it could be due to greater holding losses than average. These two investors would appear to have the same disposition effect when only that was measured, but are found to have different patterns of trading behaviour when gains and losses are measured separately.

Summers and Duxbury (2012) found that different emotions are associated with selling gains and losses respectively. If the causes of these biases are different, then the biases themselves should be different. Richards (2012) finds tentative evidence that reappraisal affects selling losses but not gains; again, this difference is not predicted in the account from prospect theory. Furthermore, Richards also finds that using stop losses reduces the disposition effect by reducing holding losses, but leaves cutting gains unaffected.

Rau (2014) finds that women have higher disposition effects. However, this difference is driven by greater holding of losses, and there is no difference in the treatment of gains.

Weber and Welfens (2007) present significant evidence that the two sides of the effect, holding losses and closing gains, should be treated as two independent biases. They include one field and one experimental study. The field study includes trades of a population of investors from online broker, and they find wide variation in disposition effect in the population of investors, as per Dhar and Zhu (2006). However, in addition, they find that cutting gains and holding losses are not correlated (after controlling for portfolio size). They also show medium-size correlations when analysing cutting gains and holding losses when measured annually.

Their experimental study consists of two tasks: one like Weber and Camerer (1998), and another using a simpler lottery task framed as timing a house sale. Participants repeated both tasks of the experiment again a month later. As with the field study, the correlation between cutting gains and holding losses was not significant. To measure stability of the biases, scores were compared across repeated rounds of the experiment, and in addition, when the experiment was repeated the following month. There were moderate but significant correlations for cutting gains, supporting its status as a stable bias, and the same was found for holding losses.

All the studies above point to gains and losses being treated differently in a way which prospect theory cannot capture. They suggest that cutting gains and holding losses may be separate biases rather than two expressions of a unitary disposition effect.

2.2 EMOTIONS, EMOTION REGULATION, AND DECISION MAKING

The next section of this literature review evaluates the wider field of how emotions affect decision making, to assess whether similar effects have been found on other decision-making biases. It also introduces emotion regulation and particularly cognitive

reappraisal, and examines research which alters decision making by using emotion regulation.

2.2.1 The impact of emotions on decision making

Emotions can affect our decision makings in many ways. This section begins by examining how emotions can affect decisions within the context of calculations of subjective utility, such as detailed in prospect theory. Emotions can substitute partly or entirely for objective calculations of value. Emotions may lead to quantity insensitive valuations, and to exacerbate the tendency to overvalue certainty in outcomes when making decisions. Surprisingly emotions can also alter the estimation of the likelihood of an outcome itself. Stepping back from the inputs into decision making calculations, emotion may act as a meta-factor in what kind of decision making process is carried out.

The section continues by discussing physiological evidence that emotions are used as a heuristic in decision-making, and may be necessary for normal levels of performance. It is not clear though whether the physiological arousal produced is the causal mechanism for this, or an artefact used for proxy measurement of emotion by researchers. In addition, it seems likely that using emotion in this way is a rather blunt tool for decision making: for example, emotions can signal the presence of high risk but not to weigh whether the risk is worth taking.

The section ends by examining findings from neuroscience. Using several experimental setups such as the ultimatum game, hyperbolic discounting and moral dilemmas, studies have shown that when emotion is driving decisions, different areas of the brain are active compared to when deliberative decision are made: roughly, the limbic brain and insula for emotions, and dorsolateral prefrontal cortex for deliberation. This gives credence to the distinction between system 1 and system 2 drawn in the literature to define different types of decision making. In addition, the anterior cingulate cortex appears to play a role in mediating conflict between the two systems.

2.2.1.1 *Emotion as a direct input in decision-making*

Schwarz (1990) sets out his “How-Do-I-Feel-About-It” heuristic for decision making. To value an outcome, we simply reflect our feelings towards it, also known as affective valence, rather perform some kind of rational analysis to arrive at an objective valuation. The latter would require us to have at least one criterion for its value, and assess how well that criterion is met, so using affective valence as an estimate for its value is much quicker.

There is a well-known theoretical distinction between two means of decision making, named system 1 and system 2 (Sloman, 2002). System 1 is characterized as rapid, automatic, and often unconscious, in contrast to system 2 which is deliberative, objective and rational. Schwarz’s model would seem to fit into system 1: by using “feelings as information” as he puts it, it is possible to make decisions much more quickly than a deliberate valuation process. However, this is at the cost of increased likelihood of a wrong decision, since feelings may not accurately reflect the objective value a decision maker should attribute to an outcome.

We can put this into wider context of prospect theory (Kahneman and Tversky (1979), a theory of normative decision making where the subjective utility of each option with i possible outcomes is assessed as:

$$\sum w(p_i)v(i)$$

where (p_i) is the probability a specific outcome will occur, v is its value to the decision maker (the value function), and w is the decision maker’s reaction to the probability (the weighting function).

Prospect theory itself does not prescribe any method of assessing value. In Schwarz’s heuristic, instead of deliberating over the factors that contribute to an outcome’s value, the decision maker simply reflects on their emotional reaction to it and takes that as the value to be used in the subsequent decision making process (Finucane *et al.* 2000).

A problem with this approach is where an emotion being felt at the time has been generated from a source other than the target of the current decision, and should be disregarded as irrelevant. For example, when people are in a positive mood, they can misinterpret this as being indicative of an emotionally favourable response towards the target, and vice versa with negative moods. Han *et al.* (2007) present ATF (appraisal-tendency framework), which predicts how and why this process of emotions 'carrying over' will occur to affect future decisions. They focus on applying this to judgements of risk and monetary value, both of which are very relevant to trading in financial markets.

There is evidence that people are sometimes aware of the distracting role emotion can play. Kelley (1973) showed that when attention is drawn to the experimental elicitation of an affective state, participants disregarded its importance in their decisions. Thus, the effect of emotion on decision making is a function of its perceived informational value: people can adjust for emotions when those emotions are consciously identified as not indicative of their valuation of an outcome. Unfortunately, background emotions are not always explicitly identified as such.

Another problem is that our emotional reactions to an outcome may not be a good estimate of the objective value which that outcome has. In other words, the correlation between emotional and objective valuation is often moderate, leading to over-or under-valuation and subsequent erroneous decisions.

Emotion can also have more subtle effects on valuation. Instead of an emotional reaction substituting entirely for the value function, it can moderate the effect that other relevant factors in valuation should have. Kahneman, *et al.* 1999, found that emotion-laden outcomes can lead to quantity insensitive decisions, where people largely respond to the affective valence of the mental image elicited. Similarly, the prospect of losing hedonic goods (those with high affective valence) can lead to greater loss aversion (Dhar and Wertenbroch 2000). They explain this as a result of higher value being placed on those goods because something like Schwarz's heuristics is being carried out to assess the

value of the prospective loss: we “feel” the loss of high affective goods more than low ones. Of course, this makes no rational sense.

2.2.1.2 Emotion as a moderating factor in decision-making

A well-known implication of prospect theory that people overvalue certainty in their estimates of utility (Tversky and Kahneman, 1992), such that w is more than 1 where p is close to zero, and less than 1 where p is close to 1. Rationally there should be no weighting at all. Rottenstreich and Hsee (2001) found that these effects are exacerbated for affect-rich outcomes, suggesting that the emotional pull of an outcome also impacts on the w term in the subjective utility calculation, as well as on v .

In addition, emotion appears to affect not just the reaction to probabilities (w), but the perception of probabilities themselves (p_i). Affect-congruent events are judged as more likely (Kahneman and Tversky 1983) and incongruent ones less likely (Mayer, Gaschke *et al.* 1992), even if the affect is elicited by non-task manipulations (and so is an irrelevant factor). There is also a strong negative correlation between the risk attached to a target and the value attached to it (Ganzach 2000), whereas no relationship would be predicted *a priori*; this is hypothesised to be caused by the emotional response to risky outcomes.

2.2.1.3 How emotions are manifested physiologically to affect decisions

Bechara *et al.* (1994) (also detailed in Damasio (1994)) claimed to have found physiological evidence for using emotions as a heuristic. Their seminal study associated damage in the ventromedial prefrontal cortex, an area involved in emotion processing, with poor performance on the Iowa Gambling Task (IGT). The evidence for the role of emotion, or lack of it, was the simultaneous failure to produce physiological arousal (measured by skin conductance response - SCR) when contemplating risky and ultimately ruinous choices. The claim is that this physiological arousal is the physical manifestation of emotional arousal in response to future outcomes: emotions function subconsciously to signal tacit knowledge about the choices not available to

consciousness, by representing this on a psychologically available one-dimensional scale (i.e. positive to negative). The one-dimensional scale allows decisions to be made by comparing different options easily. In a similar study, Werner *et al.* (2009) demonstrated that the ability to perceive viscerosensory feedback is associated with making better decisions.

2.2.1.4 Using emotions as a more successful means of decision making than rational analysis

Returning to Bechara *et al.*, a third issue is that the conflation of risky decisions with unfavourable ones. The emotional response is not signalling some nuanced and subconscious weighing up of the benefits and risks of picking certain packs, to make the objectively best decision, but simply a reaction to risky potential outcomes. The effect produced is not optimal decision making from reading the signals produced by subconscious information processing, but simply risk aversion. Since the experimental setup correlates risk avoidance with better performance it appears as if using emotions as rules of thumb could be a better strategy in general. Emotion signalling risky options is a much more basic cognitive mechanism than emotions signalling the optimal gambling strategy in a complex task.

The same group of researchers used a different experimental setup to test how subjects behave when the optimal strategy and risk aversion conflict with each other. They found that that reliance on emotions in this case leads to poorer decision making. (Shiv *et al.*, 2005). Subjects with brain damage in several areas related to emotion (the orbitofrontal cortex, the insula, and the amygdale) risked significantly more in an investing task than normal subjects who had lost money on the previous gambling round. The key difference was that \$1 was invested each round with a 50% chance of winning \$2.50, thus a rational investor should gamble every time regardless of previous results. However, only 40% of normal subjects, held back by acquired negative emotional responses, invested immediately after losing money; this compares with 80-95% for subjects with the various types of brain damage.

Interestingly, subjects who had gained in the previous round also only invested around 60% of the time, compared to 80%-95% for the brain-damaged subjects, even though it is clear that further gambling would on average lead to further gains. This avoidance of risky decisions once wealth has increased accords with the risk aversion effect detailed by prospect theory Tversky and Kahneman (1981). Whatever the reasons, a connection between emotion area damage and persistent gambling is intriguing. Those who make financial decisions chronically such as investors and traders want their decision to be unaffected by their previous gambles, and it seems emotions may somehow be an obstacle to that goal.

Prospect theory also predicts that given previous losses, people will gamble to make up those losses (i.e. become loss averse), not gamble less as found by Shiv *et al.* However, in prospect theory decisions are determined by whether decision makers are at an overall gain or loss, rather than the outcome of the previous round. Thus, future research could analyse the choices in terms of whether a subject had gained or lost money overall when they made a decision about investing, instead of only what the outcome of the previous round was. It is unclear from Shiv *et al.* what the results would be if analysis were carried out this way.

This study supports the interpretation of Bechara *et al.* that emotions were signalling risky strategies, not optimal strategies. People who lost money may have stopped gambling, despite the clear rational motivation to continue, because of the negative viscerosomatory feedback they received when contemplating gambling which was produced due to their previous losses, overriding the long-term benefit of playing. It also supports the broader point that emotion mechanisms in the brain can influence decisions about risk. However, note that the task used in Shiv *et al.* is not logically equivalent to that in Bechara *et al.*, since in the latter subjects were dealing with uncertainty (they did not know what the packs contained) whereas in the former subjects only dealt with risk (they knew the probabilities of the potential outcomes).

2.2.1.5 *The neuroscience of emotional decision making*

Brain imaging has allowed us to see which parts of the brain are active when we make “emotional” decisions, and contrast this with rational thinking; to phrase it another way, is there any neurophysiological basis to the system 1 and system 2 distinction? Several areas of decision making have been reviewed by McClure *et al.* (2007) where emotional and rational decision making come into conflict.

It appears that emotional responding leads to a different decision than rational deliberation. Sanfey *et al.* (2003) found significantly higher activity in the insula of the responder when the ultimatum game offer was rejected. At the same time, dorsolateral prefrontal cortex (dLPFC) activity was higher when playing the game but activity was not dependent on offer rejection. The insula was also associated with anger by Phillips *et al.* (1997). It suggests that activation in the insula is associated with higher levels of anger, and the higher levels of anger “override” the rational deliberations about decision making.

Sanfey *et al.* also found a correlation between activity in the anterior cingulate cortex (ACC) and the acceptance of unfair offers. They suggest this area is needed to override the strong behavioural impulses which anger gives rise to.

Hyperbolic discounting occurs when people ascribe more value to outcomes closer in time to the present; more formally, there is an inverse relation between the value of an outcome and an exponential function of the period between the decision and the outcome. McClure *et al.* (2004) found the greater limbic and paralimbic areas (involved in emotion and reward systems in the brain) showed greater activation for choices involving immediate rewards. The hypothesis produced is that when an outcome (e.g. a reward) can be taken very soon in the future, emotional responding is activated more than usual. Then just as when playing the ultimatum game, the emotional response can lead us to overlook rational assessment of an outcome’s value.

Again, reminiscent of the ultimatum game findings, the dLPFC and ACC showed greater activity during difficult choices, and activity levels in those areas correlated significantly with choosing a delayed reward. This suggests that they are necessary to resist the pull of the emotion-rich, immediate reward option.

The studies point to a separation in the brain of deliberative and emotional motivations behind behaviour. The relative levels of activation in the dLPFC, dACC, insula and limbic system often correlate closely with what decision is made. It is possible that the distinction between system 1 and system 2 is not just theoretical, but has some basis in the neurocognitive implementation of these types of decision making, at least when system 1 is characterised by using an emotional response as the basis for making decisions. System 1 here is driven by emotion-related activation in the insula and limbic system.

This hypothesis is supported by comparative phylogeny. The human cortex has expanded greatly during our evolution and is much larger than would be expected: even compared to other primates our brain is approximately three times as large as would be expected from body size (Schoenemann *et al.* 2005). The prefrontal cortex, particularly implicated in system 2, is even more disproportionate. Other animals appear to have automatic reactions and emotions to stimuli, but they lack (to various degrees) the ability to reason about novel problems and override their natural reactions. Thus, the major innovation in human cognitive ability would be to counteract the automatic response produced by system 1 and consider the output of system 2 as an alternative.

It is known that other factors can influence the strength of the effect of emotion on decision making. For example, increasing time restraints causes individual to seek cognitive shortcuts (Siemer and Reizenstein 1998), such as using emotions as a method of valuation (as described in Schwarz's heuristic above). Shiv and Fedorikhin (1999) found increasing cognitive demands has the same effect: high memory load caused an increase in the likelihood of choosing chocolate cake over fruit salad. It is reasonable to

hypothesise that memory load is not changing the emotional response to the cake, but interfering with the ability to counteract the emotional response. If so, we would expect to see high memory load time restraints decrease activity in the dACC where this conflict mediation takes place.

2.2.2 What is emotion regulation?

Given the pervasive effect which emotions can have on our decisions when experienced, it would be highly desirable to be able to control them or mitigate their effects. This is known as emotion regulation. Individuals use a variety of techniques to regulate their emotions. This section defines emotion regulation, and identifies the form of emotion regulation used in this thesis.

Koole's review of the area (2009) has the aim of developing a classification scheme for the many types of emotion regulation which exist. His taxonomy of emotion regulation sorts strategies into a 3 x 3 matrix using the emotion-generating system targeted, and the function of the regulation, as criteria.

Koole defines emotion regulation as "the set of processes whereby people seek to redirect the spontaneous flow of their emotions". Thus, this definition encompasses only how a person can leave an emotional state already existing, and excludes emotional sensitivity which relates to how a person may affect their entry into an emotional state. His definition coincides with the layman concept of how people "deal with" their emotions once they have already arisen. Also note that his definition excludes the regulation of one person's emotions by another person, for instance a parent trying to regulate their child's emotions; these instances would be intuitively excluded too. For the purposes of this paper emotion regulation will refer only to how a person regulates their own emotions.

Koole notes that emotion regulation strategies vary by the emotion-generating system they target. This targeting includes: redirecting our *attention* (Rothermund, Voss *et al.* 2008), changing *emotion-relevant knowledge* (Gross 1998), and changing the

physiological consequences of emotion (Porges 2007). Strategies can also vary by the function of the regulation, and he notes three functions: hedonic needs, specific goals, and facilitating the global personality system. Combining these two criteria gives a 3 x 3 matrix for the categorisation of emotion regulation strategies (shown in figure 2.2 below with an example for each element). Note that Koole does not preclude the existence of other targets of emotion regulation and other functions, but merely notes that his schema covers most of the common ones that have been researched. I discuss his categories briefly below.

Figure 2.2 Emotion-regulation strategies (from Koole, 2009)

Target by function classification of emotion-regulation strategies			
<i>Emotion-generating system</i>	<i>Psychological function</i>		
	<i>Need-oriented</i>	<i>Goal-oriented</i>	<i>Person-oriented</i>
Attention	Thinking pleasurable or relaxing thoughts (Langens & Mörth, 2003); Attentional avoidance (Derakshan et al., 2007)	Effortful distraction (Van Dillen & Koole, 2007); Thought suppression (Wenzlaff & Wegner, 2000)	Attentional counter-regulation (Rothermund et al., 2008); Meditation (Cahn & Polich, 2006); Mindfulness training (Brown et al., 2007)
Knowledge	Cognitive dissonance reduction (Harmon-Jones & Mills, 1999); Motivated reasoning (Kunda, 1990); Self-defence (Tesser, 2000)	Cognitive reappraisal (Gross, 1998b; Ochsner & Gross, 2008)	Expressive writing (Pennebaker, 1997); Specification of emotional experience (Neumann & Philippot, 2007); Activating stored networks of emotion knowledge (Barrett et al., 2001)
Body	Stress-induced eating (Greeno & Wing, 1994); Stress-induced affiliation (Taylor et al., 2000)	Expressive suppression (Gross, 1998a); Response exaggeration (Schmeichel et al., 2006); Venting (Bushman et al., 2001)	Controlled breathing (Philippot et al., 2002); Progressive muscle relaxation (Esch et al., 2003)

Note: Cited articles refer to relevant empirical demonstrations or literature reviews.

2.2.2.1.1 Emotion-generating systems

The first emotion-generating system which can be targeted is attention, the process by which incoming information is selected for further processing. This has been widely studied in cognitive psychology in areas other than emotion regulation, and is now

featuring within it too (Derakshan *et al.* 2007). The emotion regulation takes place by withdrawing attention from information causing an undesired emotional response; the lack of further processing from the reduced attention lessens or removes its influence on the emotional state experienced.

The second is changing emotion-relevant knowledge. Several methods can be used to achieve this; however, the one most frequently studied is cognitive reappraisal. A cognitive appraisal is where information received is used to deduce knowledge (i.e. opinions) about the world, and these appraisals may lead to an undesirable emotional response being elicited. Emotion regulation proceeds by reappraising the available information, to change its significance or meaning, such that the new appraisal no longer gives rise to the unwanted emotion.

The last system targeted is the embodiment of emotions such as: facial expressions, bodily positions, motor movements and psycho-physiological responses (Mauss and Robinson 2009). There may be some scepticism that the body is an emotion generating system rather than only the expression of emotions. However, Esch *et al.* (2003) have provided evidence of the "reversed" causal route by showing that techniques such as progressive muscle relaxation can be effective in altering emotions; other studies by Niederthal (2007), Zajonc (1998) and others have also demonstrated similar evidence. In addition, bodily intervention may be the method by which another emotion-generation system is targeted, for example eating enjoyable food can distract our attention from an undesirable emotion, or from emotion-relevant knowledge causing an undesirable emotion.

2.2.2.1.2 Functions of emotion regulation

The most intuitive of Koole's three functions of emotion regulation is hedonic regulation: the promotion of pleasure and the prevention of pain. The second function is goal-orientated regulation: when the emotion is regulated a means to some other end. Tamir *et al.* (2007) argue fear may be encouraged to promote avoidance of a stimulus. Alternatively, we may feel that some situations require an unemotional approach, in

which case all emotions may be down-regulated (Erber, Wegner *et al.* 1996). An important example of goal-orientated regulation in the literature is expressive suppression, for example someone dealing with a customer may attempt to suppress any signs of their anger to avoid revealing it to the customer (e.g. Wallace *et al.*, 2009)

Koole's final type of function is "person-orientated" regulation. Here emotion regulation is employed to "promote flexibility in personality functioning" (Rothermund, Voss *et al.* 2008) and to "promote coherence and long-term stability within the overall personality system (Baumann, Kaschel *et al.* 2005). Koole notes that compared with hedonic and goal-oriented regulation, person-oriented regulation is holistic in its focus, sensitive to context, and integrative of potentially antagonistic aspects of the personality system.

This is the least intuitive of his functions: what would it mean if the overall personality system was unstable in the long-term, or if personality functioning was inflexible? Figure 2.2 above shows some examples such as meditation, mindfulness training, and controlled breathing. A layman's term for the concept might be the promotion of mental wellbeing and the absence of neuroticism (indeed neuroticism is associated emotional reactivity and low emotional intelligence).

2.2.2.2 The impact of emotion regulation on decision making

There are two main ways that emotion regulation could affect decision making. The first is by mitigating or removing the effects that emotions have on decision making. Using emotion regulation to change or remove emotions will thus indirectly change the decisions those emotions would have affected. A second way is by the process of emotion regulation itself competing with decision making for cognitive resources. The brain has limited processing capacity and performing emotion regulation takes up some of that capacity, leading to a fall in performance for simultaneous cognitive activities. While there are many methods of emotion regulation, most research has studied the effects of either reappraisal or expressive suppression, often by comparing the two, so they will also be the focus of this section.

Heilman *et al.* (2010) compared the effects of reappraisal and suppression on risk-taking and dealing with uncertainty. Their first study involved experimentally induced fear and disgust, and instructions for participants to reappraise, suppress, or do neither.

Reappraisal, but not suppression, led to reduced levels of fear and disgust; suppression had no significant effect on either dependent measure with either fear or disgust, whereas reappraisal increased risk-taking and dealing with uncertainty for both emotions. This result was expected as fear had previously been shown to increase risk aversion (Lerner and Keltner 2001), while suppression has been poor at reducing how strongly an emotion is experienced. Behavioural suppression appears to stop the outward of expression of an emotion, but leaves the phenomenological experience and presumably its effect on decision making intact.

Their second study involved incidental emotion regulation (measured retrospectively) in response to naturally occurring positive and negative emotions (subjects were recruited and tested immediately after receiving exam results). As predicted, negative emotions reduced risk-taking while positive emotions increased it. Reappraisal (but not suppression) caused an increase in risk-taking with negative emotions, while both suppression and reappraisal were effective in negating the increase in risk-taking produced by positive emotions. The studies taken together support the view that emotion regulation can affect decision making by changing the emotions experienced; for negative emotions, reappraisal but not suppression is effective, while both are effective for positive emotions.

Sokol-Hessner *et al.* (2009) found that reappraisal was effective in reducing loss aversion. *Prima facie*, this does not involve emotions, as the instruction was to reappraise gains and losses by changing one's perspective to mimic a professional trader. However, loss aversion is hypothesised it is caused by negative emotional reactions to the prospect of a loss crystallising when closing a trade, which encourages unwarranted risks to avoid it. To test this hypothesis, Sokol-Hessner *et al.* carried out a second study which replicated the first, but also measured subjects' skin conductance

response (SCR) to their gains or losses as they played the game. The relative effect of reappraisal on loss aversion was similar to the first study; however, they also found that losses were more arousing (as measured by SCR) than gains, and the difference between an individual's arousal to loss versus gains correlated with their degree of loss aversion (about $r = 0.4$ for both perspectives taken).

In both studies, there were between-subjects differences in the extent of reduction in loss aversion when using reappraisal. When split into two groups based on whether subjects showed a significant decrease in loss aversion when reappraising, only the group that did so showed a significant decrease in the difference between SCRs to losses and gains when using reappraisal. This result suggests that loss aversion is not a cognitive error in judgement as some have argued, but a consequence of using automatic reactions to losses (versus gains) to make decisions.

However, since this study is only correlational, there remain questions about the etiology of these differences in arousal. It could be that losses cause emotions, emotions cause higher arousal, and subjects use their arousal level as a cue to make decisions, as also suggested with heart rate by Werner *et al.* (2009). Alternatively, emotions cause the loss aversive behaviour somehow while simultaneously producing higher arousal, which the experimenters have effectively used as a proxy for measuring the emotion, thus arousal would have no causal role on decisions. Additionally, since emotions were not specifically measured here, the arousal may be produced in the absence of any emotion and used as a cue for decision-making. This is the same issue found in Bechara *et al.* discussed above, which also used SCR. SCR was taken a measure of emotion without emotion being directly measured. The work was then used to justify emotions as being an integral part of decision making, though only SCR had been shown to be associated with decision making experimentally.

The study by Shiv *et al.* (2005) (discussed in part 1) provided further evidence that effective emotion regulation could increase performance where emotions have a negative

effect. On a task where risk-taking was beneficial in the long-term, subjects performed better if they had normal cognitive reasoning skills but brain damage in areas connected with emotion (orbitofrontal cortex, insula or amygdale). Normal subjects decreased gambling in response to both losing and winning money. The exact role of emotion in affecting the decision is again unclear. However, the study suggests that if emotions could be nullified using emotion regulation, functionally analogously to the neurological damage in emotion-related areas, performance could be improved when emotions are detrimental to decision making. It also, of course, has the advantage that emotions can be allowed to return and affect decisions in more propitious circumstances.

2.2.2.2.1 Emotion regulation competing for cognitive resources with decision making

Emotion regulation requires cognitive effort itself, so increases cognitive load. Wallace *et al.* (2009) found that performing emotion regulation altered performance on another simultaneous cognitive task that did not involve emotion regulation for its performance, and this effect is caused by the effect of emotion regulation on task focus. Suppression correlated negatively with performance, but reappraisal correlated positively. Both effects were statistically mediated (Shrout and Bolger 2002) by their task focus during the task, as previous work had suggested (Kanfer *et al.* 1994). When the indirect effect via task focus was removed, the correlations were no longer significant, confirming that the effect of emotion regulation was through its cognitive competition with task focus, rather than through emotional influence on performance as in the previous section. The sizes of these correlations were small ($r = -0.23$ for suppression and $r = 0.19$ for reappraisal); however, we would not expect a huge effect given the many sources of variation in subject's task performance.

These results were expected: the cognitive work required for reappraisal only needs to be done once for each piece of emotion-relevant knowledge, while suppression requires constant monitoring and suppression of expressive behaviour so will be continually competing for cognitive resources. Cognitive dissonance created during suppression (see

footnote 4) from the mismatch between internal feelings and outward behaviour could also interfere with task performance.

With the success of the initial lab study, Wallace *et al.* carried field work in a retail environment and a call centre. Suppression reduced task performance in both ($r = -0.31$ in both), whereas reappraisal had no effect in the former (though it did improve task focus, which itself improved task performance) and improved task performance in the latter ($r = 0.28$). Again, when the mediating effect of task focus was controlled for, neither strategy significantly predicted task performance. Effects sizes were larger in the call centre, a possible explanation being that call centre work involves more interaction with customers so will be more emotionally arousing; the greater amount of emotion gives more scope for emotion regulation to cause a significant difference between the groups.

Future work could examine decision in real-time, while simultaneously measuring their emotional arousal, and relating this to their emotion regulation strategies. For instance, traders and investors have a cognitively very demanding profession, with high levels of working memory required, but which is also a very emotionally arousing environment too. Traders vary widely in the methods they use to deal with their emotions: given the emphasis within trading culture of being unemotional and detached, many resort to using techniques resembling suppression, such as trying to ignore one's emotional state, or simply behavioural expression to appear unemotional to colleagues. The work of Wallace *et al.* implies will lead to lower task focus and thus lower performance.

However, a problem is the conflation of the two ways emotion regulation may affect decision: altering emotions directly or competing for cognitive resources. Wallace *et al.* avoided this problem by using a task purportedly unaffected by emotional state. This is not always possible and conflation occurs in some of the most interesting areas to study such as trading. Decisions are also prone to emotional effects (indeed much of the work

showing as much has been carried out by behavioural economists), thus any emotion regulation would be expected to have strong effects via both routes described above.

2.2.3 Conclusions about emotions, emotion regulation and decision making

In contrast with conceptions of man as a rational utility maximiser ('homo economicus'), emotions are known to affect decision making in a multitude of ways. Emotions can be used as a proxy for the value of a choice, they may bias the objective estimation of a choice's value or probability of occurring, or they may act as a meta-factor in affecting the balance between automatic and deliberate cognition, often referred to as system 1 and system 2. However, there is much potential for research in the effect of specific emotions on specific types decision making. The most basic division between emotions is between positive and negative emotions, e.g. happiness versus, fear or disgust, and most research has been carried out on this basis so far. However, it is likely that every type of emotion has specific effects on decision making.

There is evidence that not only can emotions affect decision making but that in many cases emotions, or the viscerosomatic feedback taken to be a manifestation of them, appear to be essential for it to be proficiently carried out. Damasio (1994), Bechara *et al.* (1994), and others have shown that brain damage to areas involved in emotional responses to gambling impairs performance compared to normal subjects. In addition, this coincides with the production of high skin conductance responses (SCRs) to risky decisions. Indeed, performance improved in line with increased SCR even before normal subjects were consciously aware of the strategy they were following. Werner *et al.* (2009) also provided evidence that another type of viscerosomatic response, heart-rate, can be used to improve decision-making.

It could have been that viscerosomatic feedback is signalling the output of some extraordinary and subconscious ability to work out the best option, as modelled by a normative theory such as subjective utility theory. Instead however, these results appear to be the result of the feedback signalling risky or uncertain outcomes, where

their avoidance coincidentally improves performance in the tasks used by experimenters. Where risk-taking results in improved performance such as in Shiv *et al.* (2005), brain damage to emotion-related areas of the brain actually results in better performance than normal subjects. This is presumably caused by the lower levels of risk-averse emotional reactions in the brain damaged subjects, resulting in more risk-taking. Whether the viscerosomatosensory feedback is perceived directly to make decisions, or is simply a by-product of emotion which is useful as a proxy measurement of emotion in experiments, is also unclear.

Studies in neuroscience have found that choices eliciting emotions activate different areas of the brain to ones which involve only deliberative thought. Not only that, but the strength of the activation in the emotion areas is correlated with the probability that a subject will choose the "emotional" response over the rational one. This dissociation between brain areas responsible for deliberate and automatic (in this case emotional) responses to decision making is reminiscent of the models of decision making dividing it into the rapid, heuristic based system 1, and the slow, deliberative and rational system 2.

Emotion regulation may affect decision making by changing the emotion experienced, thus indirectly mitigating the effects the emotion targeted would have had on decision making were it not regulated. Little is known about the full range of effects of emotion regulation because most research has focussed on only two regulatory strategies: behavioural suppression and reappraisal. However, where the effect of an emotion on decision making is already known, differences in effects between these two strategies have been robustly demonstrated. Both in studies using verbal instructions to subjects to reappraise their emotions, suppress them or neither; and in studies measuring subjects on their natural tendency to suppress or reappraise. Both for lab-induced emotions, and those which occurred naturally and were retrospectively measured.

Emotion regulation affecting decision making indirectly through emotions is supported in the case of both reappraisal and suppression. However, for negative emotions suppression is an ineffective strategy, and does not alter decision making. This is unsurprising since despite Koole's inclusion of body-targeting strategies, suppressing the outward expression of emotions, does seem *prima facie* an unpromising candidate for effective emotion regulation. By its nature it suppresses just one aspect of emotion expression, rather than dealing with the root cause of the emotion as reappraisal does. For positive emotions, both strategies appear effective in down regulating the emotion. Future research could look at the many other types of regulatory strategies which exist, their efficacy, and their effects on decision making, and the factors affecting both these features.

Emotion regulation can also affect decision making by competing for the cognitive resources used to make decisions. Wallace *et al.* (2009) have found good evidence that this effect occurs: good decision making (measured by work performance) was affected negatively by suppression and positively by reappraisal, and in both cases the effect was mediated by the effect of emotions on task focus. Again, there is little research into the effect of other types of emotion regulation via task focus. There may also be other ways in which emotion regulation can influence decision making other than mitigation of effect of emotions, and cognitive resource competition. What form these would take is an open question.

In conclusion, there is ample evidence that emotions do affect decision making. This strengthens the argument raised in section 2.1.6 that an account of the disposition effect based on emotions would be appropriate, given the literature in the wider field of decision making and emotions. The successful use of cognitive reappraisal (a form of emotion regulation) to affect decision making in previous research supports its use as a de-biasing intervention to change the disposition effect in this thesis.

2.3 APPROACHES TO JUDGMENT AND DECISION MAKING RESEARCH

This section critically examines paradigms in decision making research. It locates the research in this thesis as in the heuristics and biases paradigm. However, in reviewing the advantages of the naturalistic decision making paradigm, it explains why this thesis aims to improve ecological validity in the research it undertakes, and attempts to study experimental interventions in experienced investors as well as novices.

2.3.1 Two paradigms of decision making research

Normative theories of cognition, for example formal logic and probability calculus, set out how we should reason. However, it has been recognised for some time that humans do not always make decisions in accordance with normative theories (Johnson-Laird and Byrne 1991), especially under uncertainty. The two main obstacles are that the realities of real life preclude all the necessary information for doing so, which normative theories assume is available, and that humans rarely process information (complete or not) as prescribed by normative models. This can be due to external factors (e.g. lack of time) or internal ones (some facet of their psychology). Most economic models continue to assume normative decision makers at their heart in the form of expected utility theory (EUT), which holds that people act to maximise the utility they expect to experience by choosing from a variety of options with known consequences and likelihoods. It is demonstrably false that people do this; however, in aggregate behaviour does often accord with its predictions, giving some credence to its continued use. I will look here only at individual behaviour.

Instead of the prescriptions of normative theories, people often make judgements much more quickly. This is variously described as intuition, insight, heuristics, rules of thumbs, and others, in contrast to deliberation or formal analysis. The dual process model labels these two methods as system 1 and system 2. System 1 is a rapid, usually automatic method of making judgements, while system 2 is slow and conscious process, which includes working through whatever normative theory may be applicable to the problem at hand. On reflection, it is obvious that system 2 is not used for all judgements and decisions as it would be far too time-consuming.

Evidence for these two systems working in parallel comes from simultaneously held contradictory beliefs, with system 2 supplying the correct belief, and system 1 supplying the incorrect one but being automatically activated (Sloman 2002; Evans 2003). For example, the infamous Linda problem (Kahneman and Tversky 1983) demonstrates the conjunction fallacy, where participants judge $P(A + B) > P(A)$. Once the fallacy is highlighted, people can recognise their error. However, they are still unable to inhibit system 1 suggesting it. In this case we would argue system 1 is employing the representative heuristic described by Kahneman and Tversky, while system 2 is applying probability calculus. A perceptual analogy is the Muller-Lyer illusion, or indeed any optical illusion: knowing that a perception is deceptive and even measuring the lines does not prevent the illusion appearing.

2.3.1.1 The two paradigms

The paradigms of decision making differ in the conceptualisation of system 1. The heuristics and biases approach (HAB) attempts to explain deviations from normative models as the result of heuristics employed in system 1, such as the representativeness heuristic mentioned above. Where system 1 produces inaccurate beliefs, they are often referred to as cognitive illusions or biases. Though there is a positive side to the HAB research program in viewing heuristics as useful shortcuts which approximate the output of normative models, it tends to focus on biases instead. Thus, it usually portrays a negative view of human reasoning abilities as prone to errors (Kahneman and Klein 2009), though this is not a necessary conclusion from its premises and research findings. The initial work in the field (Tversky and Kahneman 1973; Tversky and Kahneman 1974) found that people are frequently unable to apply probability calculus correctly to calculate probabilities. Subsequent work has discovered a wide range of heuristics and biases, some adding to initial findings in economics-related areas such as preference reversals caused by framing (Tversky and Kahneman 1981), with others extending the approach to other areas such as social cognition (Nisbett and Ross 1980; Taylor and Brown 1994). Experimental data often comes from lab studies, and is only applied to

situations where normative models are applicable and participants responses can be compared against them.

The other paradigm is natural decision making (NDM). This takes an entirely different approach: it dispenses with normative models in all but very high-level goals (e.g. fire-fighters should aim to rescue people while avoiding dying themselves) and tries to describe how decision-making takes place in real life, particularly focussing on how experts make decisions in time-pressured, informationally noisy, high risk and high stakes environments. Clearly these situations do not give opportunities for extensive information analysis: some other form of decision-making is necessitated. Despite the pressures they face, many experts, such as firefighters, medics, and army officers, can make highly effective decisions using intuitive judgements. Though they are frequently unaware of the basis for these decisions, intuition is obviously not magic, and the judgements are somehow made using a combination of tacit knowledge and alternative decision-making processes. The primary goal of NDM is to identify what these are and so "demystify intuition" (Epstein 2010).

Though the two paradigms are loosely connected on the theme of judgment and decision-making (JDM), they have quite different focuses. HAB looks for deviations from normative theory and seeks explanations of it, while NDM studies areas where normative theories are impractical or inapplicable, and seeks to document the alternative methods which are used. This difference frames their approaches to expertise: HAB assumes that ideally everyone one will adhere to normative models, and indeed the examples normally used are not intellectually complex to work through (such as the Linda problem). HAB starts from an answer to a problem prescribed by normative theory and demonstrates that most people are susceptible to deviations from it, offering the explanation that biases are a fixed part of our psychology; some research even shows that experts make the same errors as others *despite* their greater expertise in the specific areas which they are being questioned about. NDM on the other hand starts from the premise that novices will be extremely poor at decision-making, since good decision require domain specific

knowledge which can only be gained from experience. These two paradigms are often portrayed as alternatives; however, the differences in descriptions of JDM they espouse appear to stem from the differing scope and focus of each paradigm. They do not make competing predictions that could be tested against each other, so much as talk past each other much of the time; one can accept both and treat them as complements rather than alternatives. Which paradigm is relevant will depend on the situation the judgement is made in, and the aims of the researcher.

That human reasoning is prone to biases from normative theory is interesting itself for HAB. The important fact is that normative theory gives a correct answer, and we sometimes get the wrong answer by following various shortcuts. NDM is interested in whether experts can pursue their goals in challenging environments. Normative models are practically impossible to implement, and essentially irrelevant. We could take an instrumental view of rationality – “do they achieve their goals in the context?”, but the minimal goals prescribed by instrumental rationality are assumed anyway (save lives, win battles, etc). What is interesting for NDM is people can make very good decisions despite the chaos around them.

2.3.1.2 The heuristics and biases approach (HAB)

There are two aspects to this approach, heuristics and biases, which are often treated together. However, they refer to distinct aspects of JDM. Heuristics are descriptions of the process of cognition; they can be thought of as mental ‘shortcuts’ which are often effective at reducing the mental load required to solve a problem, since humans only possess limited processing capacity and usually limited time. This is the positive agenda of the HAB paradigm referred to above. Biases are result of cognition where output is systematically suboptimal (compared to normative models); these can be the result of the operation of heuristics when the shortcuts do not work, but can have other causes.

Keren and Teigen (2005) provide an interesting analogy to these two phenomena.

Heuristics can be thought of as similar to fallacies in deductive reasoning: they are both

errors in drawing conclusions from processing information. Though heuristics are usually seen as operating automatically instead of consciously, fallacies can also be semi-automatic, where conclusions are drawn without full deliberation. For example, the *ad hominem* fallacy is irrelevant to whatever is being discussed, but we automatically tend to take it into consideration.

Biases are more akin to perceptual illusions, which like visual illusions appear in the mind effortlessly and “pop” into the mind, regardless of whether they are correct or not (Epstein, Lipson *et al.* 1992). “The mind has its illusions, like the sense of vision”, as noted as far back as 1814 (Laplace 1951). In the Linda problem, the feeling that Linda is more likely to be a feminist bank-teller than simply a bank-teller is compelling even once people have had the problem explained to them and agreed their intuition is wrong. In the background, the representativeness heuristic judges the similarity of Linda’s description to the options; the resultant feeling that the feminist bank-teller is the correct option leads the tendency to choose it.

2.3.1.2.1 Dual process functioning in HAB

As noted above, the simultaneous contradictory beliefs produced in response to these problems are seen by many as strong evidence that there must be two separate systems of human reasoning (Sloman 2002). System 1 works by the automatic operation of association in the memory, while system 2 requires effortful thought and linear reasoning. System 2 can be used to work out a problem from first principles (i.e. from normative theory); however, it is more often used to evaluate the output of system 1. Having these two systems allows us to make inferences based on the statistical structure of the environment in system 1, but also have the flexibility to monitor its output, and if necessary to disregard it altogether. Indeed Kahneman (2000) has indicated that this is how he and Tversky envisaged the heuristics and biases approach to cognitive working in practice.

The ability or tendency to use system 2 appears to be correlated with cognitive ability as Stanovich and West (2000) found significant negative correlations between cognitive

ability and susceptibility to many biases. However, in later work Stanovich and West (2008) have found that this relationship holds for some biases but not others. To paraphrase, their updated work proposes that the intervention by system 2 to avoid biases requires the ability, recognition, and cognitive capacity to carry it out. Although they do not assess the disposition effect, they do test framing and sunk-cost effects, which are related to the disposition effect, and find that neither is attenuated by increased cognitive ability.

2.3.1.3 Naturalistic decision making

NDM is about how real-life decisions are made given environmental constraints, and particularly researching how experts perform at a higher level than others. Combining the two, the core of NDM is about how experts manage the environmental constraints and make good decisions despite them. NDM, in a phrase, “puts the expert at the center of the investigation” (Rosen *et al.* 2008), and this focus on experts is a major difference with HAB. Whereas in HAB judgements are compared to a normative model, normative models are rarely even considered in NDM. How can one prescribe a normative model for a firefighter when each fire each unique – in a different building, with causes, layouts and combustible materials, and perhaps a different fire crew each time too? The same could be said for an army officer, facing an enemy whose own actions add complexity to the situation, and whose reactions to his decisions will likely vary too.

What NDM proposes is that experts perform better by acquiring tacit knowledge about their domain of expertise. This knowledge is non-transferable across domains, for example, an expert chess player does not have any advantage over other people when playing poker (Ericsson and Lehmann 1996). This knowledge also takes a long time to accumulate: normally at least ten years. Unlike susceptibility to some biases in HAB, general mental capacities are not valid predictors of expertise. Likewise, despite a general cognitive decline in older people, expertise can be maintained until at least seventy with regular practice in the specific domain of expertise (Horton, Baker *et al.* 2008).

The methodology of NDM usually involves studying experts and trying to work out how they make decisions. For example, asking experts to "think aloud" as they are working to capture their thought processes. Lab experiments are not precluded: they may be used to test ideas about identified facets of their expertise that have been identified, for example chess masters have been shown to have much better memory for chess positions from real games, but are little better at remembering random configurations of chess pieces (Chase and Simon 1973). However, NDM lab experiments do not work out what the "correct" way to approach a problem should be, and test participants for accurate reproduction of this. Quantitative research is still possible by measuring outcomes. By comparing experts' performance not to a normative model but to novices, it has been demonstrated that in some fields experts' performance is no better (Camerer and Johnson 1991)! However, in the absence of normative standards, the benchmark for performance becomes the experts themselves. Either exceptional individuals can be identified and studied, or experts can be compared to novices on a continuum of competence.

Whereas HAB sees laymen working through system 1, with better performers working in system 2, NDM has the opposite view. It models novices as attempting to use system 2, as they have no experience to use, while experts operate mainly through system 1. However, the nature of system 1 envisaged is very different between the two approaches. Instead of a series of heuristics, decisions in NDM are made using tacit knowledge of their domain which has been acquired with extensive experience. Since they have seen so many scenarios before, they are able to recognise the relevant cues within a scenario (Spilich *et al.* 1979). Over time, experts learn what is important, and this allows them to make comparisons across many other situations in that domain based on the key variables within it. Eventually they can discern the 'bigger picture' in any given situation. From this recognition, they can model what is likely to happen and thus what is the best course of action. NDM is most applicable where environments contain information but are noisy, such that it is not easily apparent what to pay attention to.

2.3.2 The implications of the two approaches for the role of emotions in financial decision making

2.3.2.1 *In Naturalistic decision making*

The first way that affect can be involved is as the conscious output of system 1 in NDM. When we think about how “good” a possible decision would be, system 1 can summate the expected consequences as valenced affect, that is, label it as attractive or aversive. Indeed Bechara *et al.* (1994) found that when prefrontal cortex damage prevented participants from attaching affect to possible consequences, they were unable to make adaptive decision using rational thought alone. Not only was affect necessary for avoiding costly losses in the game they played, but affect appeared to change their playing behaviour prior to them being able to consciously articulate what they were doing. This suggests that without affect, subconscious processing of tacit knowledge would not be able to signal what decisions to make, or to allow us to compare decisions based on a psychologically available one-dimensional scale (i.e. their valence).

A less strong position would be to claim not that affect is necessary to make decisions, but that it can be used as a cue if desired. This view is advanced by Schwarz (1990) in his “feelings as information” heuristic. Instead of weighing up all the relevant information in planning, we may simply reflect on our affective reaction to it. The problem is that affect can be generated from sources other than the subject of the current decision being considered. When people are in a positive mood, they can misinterpret this as being indicative of favourable affect towards the subject, and vice versa with negative moods. Indeed, the impact of the affect is a function of its perceived informational value; when attention was drawn to an experimentally elicited affective state, participants disregarded its importance in their judgements (Kelley 1973). However, background affect is frequently not adjusted for if it is not consciously labelled as such.

Slovic *et al.* (2007) put forward a very similar idea named the “affect heuristic”. The examples they give involve affect playing a mediating role to explain some biases

already established, such as evaluability of numbers and proportion dominance. In these cases, the causal role of affect is a possibility.

2.3.2.2 In Heuristics and biases

The HAB paradigm is still based on an amended version of normative theory, where decision makers make judgements about the value and likelihood of each outcome. These are called the value and weighting functions, i.e. "how much do I value an outcome and how likely is it to occur?". Affect can influence expected subjective utility by altering either of these terms. The value function is quite straightforward, being similar to the effects discussed for NDM. Instead of deliberating over the factors that contribute to an item's value, the decision maker simply reflects on their affective reaction to it (Finucane *et al.*, 2000). When affect is used, we become very insensitive to the numbers involved in a situation and largely respond to the affective image of one example brought to mind (Kahneman *et al.*, 1999). Similarly, the prospect of losing goods with high affective impact leads to greater loss aversion (Dhar and Wertenbroch 2000). In terms of expected subjective utility this makes some sense, since our subjective utility will suffer more from the loss of a good which makes us feel happier. However, from a rational point of view it does not, as goods which cost the same amount to replace should be valued equally.

Surprisingly, affect can also affect our judgements about likelihood. It is already known that people place more importance on probability differences close to 0 or 1 than they do for differences near 0.5. Rationally there should be no weighting at all of course – a 0.01 increase in probability should matter the same regardless of the initial probability. Rottenstreich and Hsee (2001) found that these effects are exaggerated for affect-rich outcomes. In addition, affect appears to affect the perception of probabilities themselves. Affect-congruent events are judged as more likely (Johnson and Tversky 1983) and incongruent ones less likely (Mayer, Gaschke *et al.* 1992), even if the affect is elicited by non-task manipulations (and so is an irrelevant factor).

2.4 IMPROVING EXTERNAL VALIDITY IN EXPERIMENTAL STUDIES

As discussed above, naturalistic decision making is about studying how real-life decisions are made given environmental constraints, and how experts manage environmental constraints and make good decisions despite them. This focuses on experts and real-life decisions is a major difference with the heuristics and biases approach. This section discusses previous experimental work on the disposition effect, and how this thesis seeks to improve on. This thesis aims to bridge a gap between the two paradigms, and make a contribution, by increasing the external validity of the experiments carried out in two ways: by increasing the ecological validity of the trading environment used in experiments, and using experienced traders as participants rather than naïve students.

2.4.1 Methods of studying the disposition effect

There are two main methods for researching the disposition effect, as discussed above. The first is analysis of secondary data from trading records, for example in the seminal study by Shefrin and Statman (1985), and latterly the key study of Odean (1998). The other method is experimental, as demonstrated in the classic study by Weber and Camerer (1998).

Using trading records has the advantage of directly measuring the behaviour of interest, so conclusions can be drawn about the behaviour of investors in real markets. This can establish whether a disposition effect exists in a specific country or market, and how it is affected by context (e.g. market conditions, time of year). Odean (1998) uses secondary data to replicate the finding that the disposition effect does occur in real markets, and to demonstrate that many proposed explanations of the effect cannot account for it.

This is sometimes combined with a cross sectional design (e.g. a questionnaire) which adds data about demographic and psychological traits, to test the association between them and the disposition effect. For example, Richards (2012) uses this method to produce tentative evidence that investors who engage in reappraisal more often have a lower disposition effect using this method.

The measurement of the disposition effect in secondary data is usually at the level of the individual investor. Trading records identify all the trades attributable to an individual; to calculate a disposition effect for each participant many trades of both gains and losses are required, to compare their likelihood. Some experimental studies sometimes lack this functionality due to the division of participants into gain and loss groups, or an insufficient number of trades being made by each participant, for example Lee *et al.* (2008).

The downside of using secondary data is a lack of internal validity. Internal validity is about whether the causal relationship that is claimed in a study is true (Bryman & Bell, 2015). Descriptive and observational studies may not want to make causal claims, but many do, and those that do are limited by only being able to show association between the hypothesised cause and the effect.

To get strong evidence for causal hypotheses, experiments allow much greater control over manipulation of variables of interest and thus are the 'gold standard' for producing evidence about causality. Participants can be randomly allocated to experimental groups, then inferences made about the effect of experimental conditions. This study aims to test whether applying reappraisal can reduce the disposition effect, so it is well suited to an experimental research design. Richards (2012) has produced cross-sectional evidence for this, and an experimental test is the next logical step.

An experimental approach is particularly appropriate because the aim is not merely to show that investors who use reappraisal have lower disposition effects. It is also to show that investors can be instructed to use reappraisal, and that it is an effective de-biasing intervention which can be implemented by investors who previously did not use reappraisal when trading. Richards (2012) found that people who habitually use reappraisal have some reduction in holding losses. In an experimental setup however, participants will be randomly allocated to groups, and the experimental group will be

instructed to begin using reappraisal while playing the game. So, this is closer to the situation the experiment is trying to model.

The trade-off for the increased internal validity achieved by experiments is a potential reduction in external validity. External validity is concerned with whether the results of a study will generalise to other situations, people and times (Bryman and Bell, 2015). In this case, would the results of an experimental study on the disposition effect also be found when applied to investors trading in financial markets?

One contribution of this thesis focusses on improving the external validity of experimental work, while still benefitting from the improved internal validity that experiments offer. It focusses on the threats to external validity from different situations and different people. It argues that previous experimental studies of the disposition effect have been carried out using simplified experimental protocols that are significantly dissimilar to the real trading environment. In addition, the participants used have not been representative of investors in financial markets. The studies in this thesis build on previous work by making improvements in both these areas.

2.4.2 Ecological validity of experimental studies

Ecological validity refers to the extent that the experimental setup is the same as natural settings.¹ In this case, this means whether the decision environment where the disposition effect is measured in an experiment matches the decision environment in financial markets which it is intended to generalise to, in aspects relevant to the decisions made. The more this is true, the greater confidence we have in applying inferences made from an experiment to financial markets, and that the generalisations made are valid. I now discuss some aspects of the decision environment in experimental

¹ Hammond (1998) and others have argued that ecological validity is an unhelpful misappropriation of a term that originally meant something else, and that the term "representative design" is more appropriate. Ecological validity was originally defined by Brunswik (1956) in the way Hammond favours. I find the thrust of Hammond's argument convincing. However, as Hammond notes, many prominent psychologists now use ecological validity in the more recent sense, and have done so for several decades. For clarity I adopt the modern sense here.

work which often differ from that in financial markets. These are opportunities to make a contribution by improving previous work.

2.4.2.1 Continuously changing prices and forced decision points

In financial markets prices change continuously, and decisions to hold or sell are made by investors in real-time against the background of these changing prices. Investors can buy and sell at any time they choose, as little or as often as they please. As well as buying and selling stocks, investors can also 'short-sell' stocks, which is equivalent to betting the price will fall in the way that buying stocks is a bet that the price will rise.

Experimental studies usually omit some or all these aspects of the decision environment, and constrain the options available to participants, so that they can achieve their experimental objectives in a practical manner. This may be to force participants to make decisions about gains and / or losses, or to force participants to make trading decisions in a standardised way which can be compared to other participants in the same study.

Experimental studies restrict decisions to a set number made at specific times in the experimental protocol, and the choices are limited to buying, holding and selling stocks. Of course, in financial markets there are no such restrictions no forced decisions about when to trade. Investors can buy, not buy, sell, or not sell at any time, knowing that the option will always be available later (though the price may have changed). Investors can trade as many times as they wish on financial markets. Investors can also choose to short-sell rather than buy.

Experimental studies also simplify decisions for participants by presenting them with a current and fixed price, and asking participants to make decisions only at that point based on that fixed price. This is obviously unrealistic compared with the decision environment investors face with continuously changing prices.

In addition, this approach may be more likely to produce a disposition effect since it focuses attention on the current price, and perhaps the comparison to the previous purchase price. This may encourage participants to think about their current gain or loss

on the position more than they would when trading for real. So, the disposition effect measured with this method may not be representative of the disposition effects that occur in financial markets.

So, there are concerns about a reduction in the ecological validity of the experimental setup, and whether decisions made in this setup are representative of decisions made in financial markets. The simplification of trading decisions that has facilitated studying the disposition effect in the lab is also one of the main criticisms when trying to generalise from those studies to financial markets.

2.4.2.2 Using predictive information about stock movements in trading decisions

To make decisions, investors may consider historical price information and their own personal gain or loss on the position (which is assumed to drive the disposition effect). However, they may also consider information relating to whether prices will rise or fall from their current value. Indeed, neo-classical economic theory suggests this is the *only* relevant information relating to the decision. This predictive information can come from a wide variety of sources, may change in real-time, and has uncertainty around its reliability.

In financial markets then, decisions are not just about the current price and price history, but thinking about future price movements and the uncertainty associated with them. Using both price information and predictive information, which both change continuously, is a considerably different decision environment to the ones participants are exposed to in experimental studies. So, this presents more concerns about the ecological validity of experimental studies which provide little or no predictive information. Indeed, studies that only provide price information may artificially induce a disposition effect. By asking participants to make decisions about selling stocks, by giving them no basis to make decisions except price information, they may infer that they are meant to use this information. Decisions in financial markets are not made with so little information.

2.4.2.3 Emotions during trading

Due in part to these dynamic features and context of uncertainty, emotion is often reported as being a prominent part of the experience of making trading decisions (Fenton-O’Creevy *et al.*, 2011). Physiological evidence also indicates that decision making during trading is emotionally arousing (Lo and Repin, 2002). Summers and Duxbury (2012) argue for the necessity of emotions in producing a disposition effect, based on experimental findings and self-reported emotion levels. Given the hypothesised role of emotions in the production of the disposition effect and the ability of emotion regulation to mitigate it, it seems important to make decisions as emotionally engaging as possible for the participants when conducting research in the lab.

However, an experimental setup where prices and decision points are fixed may be less emotionally arousing than the full experience of financial markets. A more engaging decision environment which is more representative of real-life trading should improve this, and improve ecological validity for testing hypotheses about emotions during trading decisions.

2.4.2.4 Method of Lee *et al.* (2008)

In the study most closely related to those reported in this thesis, (Lee *et al.*, 2008), students were told they owned stocks at one price, asked to imagine 4 weeks had passed. Then participants were given the “current” prices and asked to decide whether they wished to sell or hold their positions. Sometimes they were given fixed (and pre-determined) “analyst forecasts” indicating whether the stock was likely to increase or decrease in value. Even so, this setup is considerably less cognitively demanding than one where the prices and forecast information are continuously changing.

Stock movements were presented as random but were actually determined in advance, with half the participants being asked to make a hold/sell decision when presented with a gain, and the other half asked to do so when presented with a loss. Since participants

made decisions about only gains or only losses, the disposition effect was demonstrated by showing that the loss group was more likely to hold their stocks than the gain group.

This is a good, simple design to operationalise the variables necessary for experimental testing in a convenient manner. It allows trading behaviour to be observed in the lab within a short timescale by forcing participants to make decisions at specific times. It allows manipulation of whether participants were trading gains or losses, to facilitate making a comparison between gains and losses. It also controlled the predictive information participants received, to test hypotheses about the impact of this information on trading decisions. However, this setup lacks ecological validity when applied to how decisions are taken in financial markets, as discussed above.

There is some evidence that simplifying the decision environment too much can even remove the disposition effect. In Brown and Kagel (2009), the disposition effect did not occur when trading only one stock at a time when the optimal trading strategy was simple. They attribute this failure to their simplified experimental setup not triggering the framing effect which is presumed to be involved in causing the effect. Again, this suggests that to generalise from experiments on the disposition effect, experiments should be carried out in conditions which are as ecologically valid as possible.

2.4.3 Participants in experimental research

Most experimental studies use students as participants. As with the concerns about ecological validity discussed above, this raises issues about the external validity of these studies. Specifically, the issue is whether experiments carried out on one set of people can be generalised to another set of people. These studies are usually motivated by a desire to give some insight into why the disposition effect is displayed by investors in real trading. So, whether explicitly stated or not, it is assumed that a sample of naïve subjects are a good model for experienced investors and make decisions in the same way as financial investors. For example Lee *et al.* (2008) used undergraduate business

students as participants to test the disposition effect and the effect of reappraisal on the disposition effect.

In the case of this thesis, I address not only the assumption that students' decisions will be representative of investors' decisions, but also the assumption that students will respond to cognitive reappraisal in the same manner as experts. It may be the case that cognitive reappraisal is only effective on naïve participants, unfamiliar with technical trading techniques.

There are many reasons for thinking that the decision making of naïve subjects may differ from experienced investors. Many studies have produced evidence that demographic characteristics can predict the level amount of disposition effect displayed by investors. Both experience (Seru *et al.*, 2010) and age (Dhar and Zhu, 2006) have been shown to reduce the disposition effect displayed.

Investors have knowledge of common stock-trading techniques and strategies. They have personal experience of trading, and retail investors have experience of actively trade their own capital. Not only this, but most investors are aware of the disposition effect itself, and that it is not a good pattern of investing. Advice such as 'all big losses begin as little losses', 'sell at the first loss', and 'sell losers and ride winners' all implore investors to implement the opposite of the disposition effect.

How precisely reappraisal changes emotions is not known, and would require a full account of how emotions are produced by conscious thoughts, which of course we don't have. However, in making claims about reappraisal it seems better *ceteris paribus* to use participants who are as similar psychologically as possible to the population the results are being generalised to. Student samples are far from ideal for representing financial market investors.

2.4.3.1 Practical issues with participant samples

The reason that studies typically use students is clear. It is much easier to recruit a sample of students to take part in a study than a group of retail investors. Student

participants bring the benefits of large numbers, on site availability, and plentiful free time. They can typically be persuaded to participate with the promise of small amounts of money. In some institutions, participants even receive credit towards their degrees in exchange for their time. (Given the strong motivation many students have to avoid additional work, I understand this can be very appealing). Of course, these are all convincing *practical* benefits to taking such an approach, but the trade-off is greater caveats against the external validity of the studies.

Retail investors on the other hand have many characteristics which make them difficult to use as participants. They are unlikely to be available in person, and difficult to find in large numbers. They are unlikely to be persuaded to spend their time taking part in long experimental protocols for nominal sums of money. Indeed, the contrast with other uses of their time is particularly apparent for them – they could use the same amount of time trying to make money trading their own portfolio.

The use of student participants in experimental psychology, and other research which adopts its methods such as behavioural economics, is very widespread. It is an accepted compromise in the process of carrying out experimental research; thus, it is not usually seen as a weakness of research, to the extent that using such participants makes a piece of research invalid. However, this thesis argues that using representative participants is still an important improvement on such studies, which forms part of the contribution of this thesis. To make improvements in the external validity of testing reappraisal, this thesis will attempt to use participants who represent the same people that results will be generalised too.

2.4.3.2 Heterogeneous adult samples

If expert participants cannot be recruited, a smaller improvement in the representativeness of participants is to use a sample which comprises a range of adults. Using only students means that the participant sampling is confined to a narrow age range, which is more homogenous than the wider population. Importantly for this

research, it will not match the age distribution of investors either, which is significantly older than a student population profile. As noted above, age has been found to impact the disposition effect displayed (Dhar and Zhu, 2006).

Students will also tend to be more clustered than the population in a whole in range of psychometric variables, such as cognitive ability. If participants are drawn from only one institution, which is often the case, this clustering will even greater than within the wider student population. In fact, students are often drawn from a single degree program, increasing the clustering further still. Therefore, where a specific participant sample is not possible, this thesis will attempt to use heterogeneous adult samples rather than student samples.

2.5 CONCLUSIONS AND RESEARCH QUESTIONS

The literature reviewed above does not lead linearly to the research questions, so the purpose of this section is to map out what motivates the research questions by drawing together material from the literature review. Material covered earlier is briefly explained and referenced, since the aim is not to repeat the material above but to develop the logic leading to each research question and its associated hypotheses. Methods used to test the hypotheses will be described in more detail in chapter 3, and further detail is given on the individual tests for each hypothesis in chapters 4-7, which report the results of the studies carried out.

2.5.1 Research motivations shared by all 4 research questions

Although all four research questions have different focuses, they all address the disposition effect in some way. This section sets out three research motivations related to the disposition effect which apply to all 4 research questions. These are: the pervasive nature of the disposition effect and its economics effects; its status as an error or bias in decision making; and increasing the external validity of experimental studies used to research it.

To establish why the disposition effect is an interesting topic to research, this chapter began by explaining in section 2.1 what the disposition is, arguing why it is important, and reviewing some field studies from the voluminous literature on the disposition effect. In brief, the disposition effect is a very robust phenomenon in financial markets, having been demonstrated many times across a wide range of countries, market types and investor classes. It is important to investors primarily because investors who trade with a disposition effect tend to perform worse than those who don't (e.g. Odean, 1998).

It is interesting for researchers in judgement and decision making because it is a persistent deviation from normatively rational models of decision making, which seek to maximize expected returns. The disposition effect contravenes basic economic assumptions, for example Fama (1970) argues that stock markets are efficient, whereas investors with a disposition effect act as if markets are not. Despite this, rational explanations for disposition effect have been proposed, such as stock price mean reversion, portfolio rebalancing, and minimizing trading costs (discussed in more detail in section 2.1.2).

However, Odean (1998) tested and demonstrated that trading based on mean reversion does not work in financial markets: stocks which investors sold performed better than stocks which they held. Furthermore, Weber and Camerer (1998) showed experimentally that participants still had a disposition effect, even when they were explicitly given information which ruled out mean reversion as a strategy. Odean (1998) also conclusively refuted portfolio balancing and trading costs as explanations for the disposition effect. Since there appear to be no rational explanations for the effect, researchers have proposed psychological explanations instead, and this thesis builds on some of this work.

The last shared motivation, the desire to increase external validity, comes from contrasting the approaches of the two main research paradigms in judgment and decision making (discussed in more detail in section 2.3). The heuristics and biases

approach looks for biases in human reasoning abilities that arise because of our reliance on heuristics when making decisions (Kahneman and Klein 2009). The study of the disposition effect is very much in this tradition: investors hold losses too long because they rely on the heuristic of not selling at a loss, and don't hold gains long enough because they rely on the heuristic of 'banking' (i.e. selling) gains before prices can fall. However, the alternative paradigm, naturalistic decision making, is relevant too. This approach focusses on how experts make decisions in (informationally) noisy environments. Trading is certainly a noisy environment, where experts have a significant advantage over novices, so this paradigm can also be applied. A weakness of much previous experimental research on the disposition effect is that it neglects this expertise aspect and attempts to study biases in a way which sacrifices the external validity of the research: this threatens whether its findings will generalise to real-world contexts. So, this thesis attempts to increase the external validity of the studies it contains by addressing two issues: the ecological validity of the experimental setup, and using naïve rather than experienced participants. These issues were first discussed in section 2.4 above.

Lab experiments allow much greater researcher control of what is being measured, and therefore typically achieve high internal validity of their conclusions; however, using an artificially simple decision environment can reduce the ecological validity of the experimental setup (ecological validity refers to whether an experimental setup has the same features as the real-world environment which it is studying). For example, real-life trading involves continuously changing prices, and investors are free to buy or sell as many stocks as they wish, at any time they choose. However, to simplify data analysis in experimental studies, and make one participant's decisions directly comparable with another's, experimental studies often use a 'forced choice' method for decisions, where participants must make decisions at fixed and pre-determined times, and have limited trading options to choose from. This is the type of method used in Lee *et al.* (2008), a study which this thesis specifically attempts to build on.

Another limitation of experimental work on the disposition effect is the omission or reduced complexity of predictive information for participants about future price movements. In real-world trading, there is a huge amount of information which could be used to inform an investor's view of likely price movements so affect their selling decisions, and this information could change at any time. Often this aspect of the decision-making process is absent altogether. In other studies, such as Lee *et al.* (2008), such information is provided but it is simple and fixed.

These simplifications reduce the ecological validity of the studies, and in doing so, make the decisions made by participants much less cognitively demanding. So, the decisions made in these experimental studies may not be accurate reflections of the type of decisions made by real-world traders. This thesis improves these limitations of experimental work on the disposition effect by using a realistic trading simulation (the two-index game), which measures trading behaviour in the lab with greater ecological validity. Chapter 3 explains how this instrument improves on previous methods of measuring the disposition effect in the lab. This game is used in all three studies which produce data for the four research questions, to its impact is relevant to all four questions.

There is evidence that emotions can be a prominent part of the trading experience, but it is also possible that participants do not find making decisions in simplified experimental setups as emotionally engaging as real-life trading. Research questions 3 and 4 are concerned with the role of emotions in the disposition effect; the increase in ecological validity may be beneficial for these research questions not only when measuring the disposition effect, but also when eliciting emotion responses in an experimental setting.

The other issue this thesis addresses is the use of naïve rather than experienced participants. Naïve participants are easier to recruit than experts, but the use of unrepresentative participants can also weaken generalisation of the results. Studies on

the disposition effect which want to generalise results to real-life implicitly assume that naïve participants are a good model for experienced investors. Again, this was the case in Lee *et al.* (2008), which this thesis seeks to build on. To improve on this, two of the three studies presented here use retail investors instead of student samples: these provide the bulk of the data in answering questions 1 and 2, and all the data for question 3.

2.5.2 First research question

Although the disposition effect is widely studied in the field, this research doesn't allow us to distinguish between whether the disposition effect is simply a 'state' phenomenon, being simply a product of the environment where it is observed, or whether it has a trait-like dimension. What is meant by trait-like is that the bias shows reliable differences between investors, so that there is variation across the population, but that the bias is relatively stable within each investor, for example when measured across a range of contexts and across time. If the bias has a trait-like dimension, an investor with a high disposition effect in one environment would be more likely to display a high level in another environment, or when measured again at a later time. Unfortunately, field studies cannot show this directly because they only measure investors expressing a disposition effect in one context.

A common contemporary explanation of the disposition effect is based on the application of prospect theory, which was presented in section 2.1.3. Although research questions 3 and 4 and are more focused on the actual causation of the disposition effect, this explanation also has implications for the trait-like nature. It proposes that the disposition effect emerges from the shape of an individual's value function. Since this value function differs between individuals (Schunk and Betsch, 2006), the disposition effect individuals express should also differ. So, this is a theoretical argument for the disposition effect having a trait-like character.

There is also empirical evidence that the disposition effect has a trait-like character, first discussed in section 2.1.4. Many studies have found variation between individuals in financial markets, for example, Dhar and Zhu (2006); however, these studies do not establish that the effect is also stable within individuals over time and situations. Building on this, Seru *et al.* (2010) do provide strong evidence for stability over time within investors (though only with one sample and in one market). They show that over a nine-year period, while the average disposition effect in a sample of investors decreases, most of the change in disposition effect within the trading population is attributable to investors with a high disposition effect ceasing to trade. The disposition effect of each individual changes little over this period, despite a long time to adapt their trading style. This both demonstrates stability over time, and suggests that the disposition effects of individual investors may not change because it has a trait-like nature. Other researchers have found correlations between the disposition effect and other stable traits, which again suggests it may have a trait-like nature too. For example, Kadous *et al.* (2014) have shown correlations with aspects of self-esteem, and Richards (2012) has found correlations with cognitive style.

Building on this prior work, this thesis aims to provide direct evidence for trait-like characteristics of the disposition effect. First, the intra-individual stability of the disposition effect is assessed, by measuring it repeatedly. Then, convergent validity across different measures of the disposition effect is tested. Convergent validity means that variables which measure the same underlying factor are associated with each other when measured. To establish this, convergent validity is tested using three types of data: field data, experimental data, and survey data. These aims lead to the first research question and its hypotheses:

RQ 1 - Does the disposition effect have trait-like characteristics?

H 1.1 - The disposition effect will show intra-individual stability over time

H 1.2 - The disposition effect will show convergent validity across multiple measures

As noted above, the research question not only investigates the trait-like characteristics of the disposition effect, but also does so while improving the external validity of the experimental data supporting the hypotheses.

2.5.3 Second research question

The second research question addresses similar ground to the first question, and investigates the trait-like characteristics of trading biases. Rather than the disposition effect itself though, it is concerned with the two 'sides' of the disposition effect: cutting gains and holding losses. Do they also have trait-like characteristics: are they also stable behavioural tendencies of investors, rather than simply a product of the particular environment in which decisions are being made?

RQ 2 - Do cutting gains and holding losses have trait-like characteristics?

As mentioned above, a common contemporary explanation of the disposition effect is based on the application of prospect theory, and was presented in section 2.1.3, and evidence which challenges this account was discussed in section 2.1.6. In doing so, many studies point to the role that emotions could play in producing the bias, and how the behaviour towards gains and losses may be underpinned by different emotions. For example, Kaustia finds that investors often treat positions categorically, rather than being sensitive to their relative size. A loss, of any size, is much less likely to be sold; not only does this contradict a prospect theory account, but it could suggest that the disposition effect is specifically related to responses to losses. If this was the case, it would be incorrect to treat both cutting gains and holding losses as two expressions of a unitary disposition effect bias.

This possibility was developed further in section 2.1.7. For example, Summers and Duxbury demonstrate that different emotions are associated with the two sides of the effect, and Richards (2012) found tentative evidence that reappraisal affects losses but not gains. These findings suggest that different emotions may be responsible for the two 'sides', which implies that they are independent biases.

Weber and Welfens (2007) have carried out promising work in this area, by demonstrating the separation of the two biases using two experimental tasks. However, while Weber and Welfens used a trading task and a housing task, both were relatively simple lab experiments. In particular, their 'housing task' was simply a series of gambles given a housing theme, and so is very artificial. This thesis expands and improves on this study by measuring the disposition effect with greater ecological validity, and by comparing the disposition effect using a much more diverse range of methods.

In this thesis, convergent validity will be tested between data from a realistic trading game, and a self-report scale. The trading game has greater ecological validity (discussed in further detail in chapter 3); therefore, using data from the game strengthens the evidence for the research question. The game and scale are also very different from each other, so the demonstration of convergent validity is more forceful. Finally, retail investors are used as participants rather than students, again creating more powerful data to address the research question.

Like the first research question, evidence for intra-individual stability and convergent validity of the two biases is assessed, which leads to the first four hypotheses associated with this research question:

H 2.1 - Cutting gains will show intra-individual stability

H 2.2 - Holding losses will show intra-individual stability

H 2.3 - There will be convergent validity for cutting gains between a realistic trading simulation and a scale

H 2.4 - There will be convergent validity for holding losses between a realistic trading simulation and a scale

In addition to demonstrating intra-individual stability and convergent validity, additional hypotheses which can be tested. Since cutting gains and holding losses are suspected to be independent biases, they can be tested against each other to provide evidence for discriminant validity. Discriminant validity is the opposite of convergent validity: that

variables which are driven by the different underlying factors will not be associated with each other. Cutting gains and holding losses are compared against each other three ways: with data for both from the trading game, with data for both from the scale, and finally crossing data for cutting gains from the game and holding losses from the scale, and vice versa. This leads to a further 3 hypotheses:

H 2.5 - There will be discriminant validity between cutting gains and holding losses in a realistic trading simulation

H 2.6 - There will be discriminant validity between cutting gains and holding losses in the scale

H 2.7 - There will be discriminant validity between cutting gains and holding losses, between a realistic trading simulation and a scale

2.5.4 Third research question

The third research question relates to explanations of the disposition effect, argues against an explanation based on prospect theory, and explores an alternative based on emotions during trading. In addition, it also examines how reappraisal may affect expression of the disposition effect through its effects on emotions, and also affect its constituent biases (cutting gains and holding losses), which were discussed in more detail above in section 2.5.3, in the conclusions for the second research question.

A common current explanation of the disposition effect, based on prospect theory. This account applies prospect theory (Kahneman and Tversky, 1979) to the position of an investor deciding whether to sell or hold a stock they own. First, an investor must compare their position to a reference point, usually the purchase price. This creates a mental account of their position as a gain or loss. Second, an investor considers possible outcomes if they choose to sell or hold their gain (or loss). Third, these outcomes are converted from a monetary value into subjective expected utility, using the value function. Lastly, the subjective expected utility of holding a gain (or loss) is compared with the utility from selling it. The value function is curved, flattening over time in both

the gain and loss domains. However, this has the opposite effect in gain and loss domains. This curvature results in expected utility being greater when selling (rather than holding) a gain, whereas expected utility is greater when holding (rather than selling) a loss. The combination of these entails an investor is more likely to hold a loss than a gain, which is observed in the disposition effect. This account was presented in more detail in section 2.1.3, and its underlying assumptions discussed in section 2.1.5.

Section 2.1.6 proceeded to question whether an explanation based on prospect theory could account for recent empirical findings. For example, Kaustia (2010) appears to find that a disposition effect is better modelled as an 'all-or-nothing' response to holding a loss, rather than being driven by the value function as prospect theory suggests. The fact that a position is a loss means it is treated qualitatively different from a break-even position or a gain. Lehenkari (2012) finds that not being responsible for a loss reduces the disposition effect. Again, prospect theory has no role for responsibility: decisions should be driven by expected outcomes and their conversion into subjective utility by the value function. Whether someone is responsible or feels responsible should have no effect on these factors.

Building on both these findings experimentally, Summers and Duxbury (2012) found that responsibility for positions was necessary for a disposition effect. Furthermore, responsibility for losses produces an increase in feelings of regret. Again, this is not consistent with a prospect theory account. Rau (2015) also supports this: investing as a pair rather than individuals produced greater feelings of regret, and coincided with greater levels of the disposition effect.

So, in findings evidence against prospect, many of these studies suggest indirectly or indirectly that emotions are involved whether a disposition effect is expressed, and particularly that negative emotional responses towards losses lead to more holding of losses. Situations which give rise to certain emotions will lead to a disposition effect being produced. In contrast, situations or interventions which reduce or eliminate the

experience of emotions (for example when investors do not feel responsible for their trading positions) mean the disposition effect will be reduced or eliminated.

Lee *et al.* (2008) tested a different method of disrupting the disposition effect. They used a cognitive reappraisal instruction (a form of emotion regulation) and asked participants to imagine they were trading for someone else, and found this also reduced the disposition effect. Though they did not directly measure emotions, an explanation is that cognitive reappraisal can reduce the emotions associated with trading with a disposition effect. They speculated that by imagining trading for someone else, participants felt more distant from the situation, and did not experience the emotions associated with trading as strongly. Supporting this, Richards (2012) found tentative evidence that using cognitive reappraisal decreases the disposition effect by reducing the holding of losses.

As background to this alternative explanation of the disposition effect, this literature review has also reviewed the wider literature of emotions, emotion regulation and decision making, in section 2.2. Section 2.2.1 looks at how emotions have been found to affect decision making, and finds that the effect of emotions is widespread: this makes an emotion-based explanation of the disposition effect plausible. For example, dual process theory has modelled decision making as comprising two systems, system 1 and system 2 (Sloman, 2002), contrasting rapid and often automatic reasoning against deliberative and objective reasoning. Emotions affecting decision making are a possible example of system 1 in action: they provide a rapid and psychologically salient heuristic on which to base decisions (Schwarz, 1990).

Emotions during trading could produce the disposition effect in a similar way. The prospect of selling a loss could produce negative emotions. These emotions lead to system 1 reasoning: investors are motivated to avoid these unpleasant emotions, and some chose to do so by not selling the loss. In other words, they prioritise hedonic goals over economic ones: by treating a paper loss as temporary they delay the emotional pain of selling the loss, hoping that the loss will reverse. This would explain the large

increase in the probability of selling Kaustia observed when a position moves from a small loss to a small gain. A small change in price produces a large increase in selling, which can be explained by a large change in the emotions experienced when considering selling it.

Section 2.3.2.2 discusses the role of emotions in the heuristics and biases paradigm. Previous research has shown that when using system 1, affect can cause people to become insensitive to the numbers involved in decision and simply respond in line with the affect experienced. This appears to be a good explanation for the findings of Kaustia, for example.

Richards (2012) also supports a potential explanation based on the balance of system 1 and system 2 when making trading decisions. He found that investors with higher reliance on system 1 have higher disposition effects, and tentative evidence that investors with higher reliance on system 2 have lower disposition effects. A higher reliance on system 2 could moderate the impact of system 1 by allowing investors to override the effect of negative emotions.

Emotion regulation is the main intervention used in this thesis, and is a key part of research questions 3 and 4. It is introduced in section 2.2.2. If the hypothesis that emotions affect the disposition effect is accepted, there are a variety of types of emotion regulation which could be used to target it experimentally. Koole (2009) has many categorised different types of emotions regulation based on psychological function and the emotion-generating system targeted. However, cognitive reappraisal is the form chosen here: Lee *et al.* (2008) used cognitive appraisal as an intervention to affect the disposition effect, which this thesis builds on, and in addition cognitive reappraisal has also been used to successfully mitigate other decision-making biases. For example, Sokol-Hessner *et al.* (2009) found that cognitive reappraisal reduced loss aversion. In fact, not only did cognitive reappraisal reduce loss aversion, but this reduction was correlated with a reduction in physiological markers of negative emotion. This strongly supports the

argument that emotions during trading could underpin the disposition effect, and that cognitive reappraisal is a promising intervention to target these emotions.

So, the third research question addresses whether cognitive reappraisal can reduce the disposition effect. Lee *et al.* (2008) have previously shown this, but their demonstration lacked external validity. This thesis tests whether their result still holds when external validity is improved, leading to the third research question:

RQ 3 - Does cognitive reappraisal affect the disposition effect and its constituent biases, when tested in experienced traders under conditions of greater external validity?

Note this increase in external validity is relevant to all four research questions, as discussed above in section 2.5.1 earlier. However, it is most important for the third research question since the increase in external validity is a key motivation for building on the previous work of Lee *et al.* (2008).

The first hypothesis for this research question establishes that a disposition effect is expressed in the study, since without it the intervention cannot reduce the disposition effect.

H 3.1 - Investors will show a disposition effect in a realistic trading simulation

The second hypothesis aims to replicate the result of Lee *et al.* (2008) but with greater external validity as discussed above.

H 3.2 - Cognitive reappraisal will reduce the disposition effect

The final hypothesis builds on the finding of Lee *et al.*, together with other evidence discussed above pointing to the role of behaviour towards losses in driving the disposition effect, and changing them using emotion regulation. So, the third hypothesis tests whether cognitive reappraisal produces its effect by reducing holding losses specifically.

H 3.3 - Cognitive reappraisal will reduce holding losses but not affect cutting gains

2.5.5 Fourth research question

The last research question follows directly from the third research question, and the motivations leading to the third research question also motivate the fourth research question. The exception to this is improving external validity by using retail investors as participants: to allow more complicated experimental procedures, it was necessary to use lay participants rather than a further sample of retail investors. However, external validity is still improved by using the realistic trading simulation.

RQ 4 - Does cognitive reappraisal affect the disposition effect and its constituent biases, by changing emotions during trading, when tested in novices under conditions of greater external validity?

Since many aims were shared between the third and fourth research questions, the first three hypotheses mirrored those for the third research question:

H 4.1 - Novices will show a disposition effect in a realistic trading simulation

H 4.2 - Cognitive reappraisal will reduce the disposition effect

H 4.3 - Cognitive reappraisal will reduce holding losses but not affect cutting gains

However, the fourth research question goes further, since it aims to directly test the involvement of emotions in the disposition effect and the effect of cognitive reappraisal on emotions and the disposition effect. So, the fourth hypothesis tests that cognitive reappraisal is having the expected effect on emotions:

H 4.4 - Cognitive reappraisal will reduce negative emotions experienced during trading

The last hypothesis tests the proposed mechanism for how cognitive reappraisal reduces the disposition effect. Since higher negative emotions are expected to lead to a higher disposition effect, reducing them should also reduce the disposition effect. The final hypothesis tests whether this causal chain is supported:

H 4.5 - Changes in emotions during trading will mediate the effect of reappraisal

3 RESEARCH PHILOSOPHY AND METHODS

The previous chapter has discussed literature relevant to the research questions in this thesis. This chapter begins discussing the research philosophy that underlies this thesis, and justifies the methods adopted given that philosophy. Research philosophy comprises the philosophical positions a researcher takes about the nature of reality (ontology), the kind of knowledge we can hold about reality and how to generate that knowledge (epistemology). Three broad research philosophies will be discussed: subjectivism, positivism, and realism. This thesis adopts realism. This is the mainstream position taken in psychological science and experimental behavioural finance, the fields this thesis contributes to.

Chapters 4-7 include discussions of methods which relate specifically to the results in those chapters. However, the second part of this the chapter discusses some methods which are shared by all three studies in this thesis (the Milan, London and OU studies). This includes a discussion of the realistic trading simulation (two-index game), and how it improves ecological validity.

Also, some methods which differ across studies are dealt with together here and compared across studies, rather than have short standalone sections in later chapters. This includes participants, protocol and participant incentives.

3.1 RESEARCH PHILOSOPHY

There are, very broadly, two main schools of thought in research: subjectivism and empiricism. They differ, again in very broad terms, in their beliefs about whether research in the social sciences should emulate the approach taken in the natural sciences (Bryman & Bell, 2015). Each has their own different associated beliefs about ontology and epistemology. Ontology deals with beliefs about the nature of reality. What exists in the world, for researchers to study? Epistemology concerns our theory of knowledge,

and how we obtain that knowledge. What kind of knowledge is an acceptable product of the research process?

Note that subjectivism is used here as a general term for many related research philosophies. This account is only a brief treatment, so only seeks to sketch out what they have in common, to demonstrate that they are inappropriate for the research carried out here.

Subjectivism typically holds a constructionist ontology. Constructionism believes that reality consists only of the perceptions and interpretations of social actors, who negotiate the meaning of their actions and environment. Reality is socially constructed within the mind of each person: this does not mean that each person has their own interpretation of an underlying objective reality, but that reality *is* those interpretations. Each person constructs their own reality through their interactions with and interpretation of their social environment.

In contrast, empiricism holds an objectivist ontology like the natural sciences. Reality is external to social actors, and it is beyond the perceptions of the individuals that are involved in studying it (Bryman & Bell, 2015). The purpose of research is to uncover and understand this reality.

Subjectivism usually has an interpretivist epistemology. Interpretivism sees humans as fundamentally different to the topics studied by the natural sciences, so the methods used and knowledge gained are fundamentally different to those in the natural sciences. It attempts to study human behaviour by understanding how people interpret the world around them. Its goal is to understand the concepts and meanings used by people, and how they shape their reality. These are elucidated by penetrating the "frames of meaning", which social actors draw upon when they construct their reality. The researcher's role is to discover what these socially constructed meanings are, then describe them in academic language (Bryman & Bell, 2015).

On the other hand, empiricism's epistemology holds that since there is an external reality (following objectivism) research should proceed by measuring it. Knowledge is obtained by gathering data about this external reality through our senses, to form theories or test hypotheses about it. Usually it favours deductive theory, where theory is used to form hypotheses, which are tested using these data. Research should be carried out objectively, simply observing the data and drawing conclusions from it.

Empiricism corresponds with the philosophy usually associated with natural sciences. A popular view is that "science is a structure based on facts" (Davies, 1968); however, an epistemology is more than this, because it needs an account of what constitute facts. Empiricism takes the view that facts are generated by objective observation of the external world. Interpretivism opposes this and holds that facts are constructed by people: facts cannot be simply observed, but can only be understood by engaging with how each person interprets their reality.

This thesis adopts an empiricist approach. Trading biases are conceived as stable patterns of behaviour that people objectively possess, similar to personality traits. One purpose of the research is to build evidence that these tendencies do indeed exist and affect trading behaviour. This demonstrates an objectivist approach to ontology.

Trading behaviour is measured directly from trading decisions made within experiments, and these measurements are considered to be objective facts. In contrast, a subjectivist approach might ask participants about their trading behaviour or their interpretation of it. This thesis does not seek to interrogate people's understanding of their disposition effect, but is interested in measuring their trading behaviour and making objective comparisons between participants and between different experimental conditions.

The effect of cognitive reappraisal on trading is calculated objectively, as changes in measurable variables that independent observers will always calculate identically.

Changes in trading behaviour do not depend on whether participants perceive their bias as having been reduced. Indeed, participants are not asked whether they think they

have a disposition effect, to avoid prejudicing the decisions they make in the game. Nor are they asked whether they think their disposition effect has been reduced after the intervention of reappraisal. Again, the approach taken here is clearly objectivist. What matters is the direct observation of trading patterns, not how participants perceive or interpret their trading patterns.

3.1.1.1 Positivism versus realism

Although the research philosophy adopted is empiricism, there are different approaches within empiricism. I will briefly distinguish two main schools of thoughts, positivism and realism, and explain why realism is adopted.

The distinction between positivism and realism is not always clear. The term positivism is used in a variety of ways, and is sometimes used as a synonym for what I have termed here as empiricism. Many writers criticise "positivism"; however, it is often not clear whether they mean to refer to positivism as set out below, or are critical of any methodology which resembles a scientific approach (Bryman and Bell, 2015). Similarly, researchers may identify as positivist, when meaning that their research broadly follows a natural sciences approach such as using quantitative methods to measure phenomena and test hypotheses about them. I draw a distinction between positivism and realism here below based on the position of the researcher towards unobservable theoretical entities.

Strictly speaking, positivism is concerned with describing the relationship between observable objects and events. Knowledge is gained by generalizing from our observations. Positivism holds that phenomena which cannot be observed are not the proper objects of scientific research. It rejects the aim of producing a description of unobservable phenomena, and positivist methods do not seek to explain behaviour (Outhwaite, 1987).

This kind of approach is common with economics, and particularly econometrics.

Econometrics is not concerned with the actual processes involved when individuals make

decisions. It is usually not interested in establishing the realism of the theories, i.e. whether theories are an accurate description of reality, but only whether they predict the outcomes they should. Can a theory accurately predict whether X follows Y? This is the kind of approach argued for by Friedman (1953): all that matters about a theory is its predictive power. The researcher attempts only to specific quantitative relationships between observable phenomena. This approach applied to the disposition effect would define it, study the conditions under which it occurs, and perhaps under what conditions it remained constant or changed. However, it would not attempt to explain why this behaviour occurred.

Realism differs from positivism in that it holds that unobservable phenomena are also appropriate objects of research. We can gain knowledge about them by testing hypotheses using observable phenomena connected to them. The aim of realism is to explain observable phenomena in terms of the underlying real objects and mechanisms which produce them. Its epistemology is that we gain knowledge by building models of structures and objects which can account for the observable phenomena (Bryman and Bell, 2015). (Constructive empiricism (van Fraassen, 1980) takes a similar approach, but without the principle that these models are a true description of reality).

In the first two research questions, this thesis seeks to establish that the disposition effect, cutting gains and holding losses have trait-like characteristics, which can be studied through their expression in observable trading behaviour. In positivist research, the trading behaviour itself would be the subject of interest. Instead, here the trading behaviour is the expression of underlying behavioural tendencies. The purpose of this research is to gain knowledge about the causes of trading behaviour by building models of behavioural tendencies which account for the observed trading behaviour.

These behavioural tendencies are not used as intervening variables, but as trait-like characteristics that people possess (MacCorquodale and Meehl, 1948). They are assumed to really exist and are a cause of trading behaviour. This view of the production

of behaviour is analogous to personality traits. Actual behaviour varies depending on a person's current mental state and surrounding environment, but personality traits are real psychological phenomena that produce differences in behaviour between people. Therefore, this thesis adopts a realist position regarding behavioural tendencies in trading.

In the second two research questions, the expression of trading behaviour is manipulated using emotion regulation. This tests how observable trading behaviour is changed by different experimental conditions. However, this is associated with a model where emotions experienced during trading play a casual role in producing these trading behaviours. It is proposed that the effect of cognitive reappraisal is mediated by these emotions: a reduction in emotions experienced is the mechanism by which cognitive reappraisal is hypothesised to function. The claim is that changes in unobserved emotions are causally responsible for changes in trading behaviour, and the OU study attempts to indirectly measure emotions to test this hypothesis. Again, this is a realist perspective.

As mentioned above, the approach adopted here is also adopted within mainstream psychological science. This "seeks to discover, describe, and explain psychological phenomena and processes through the logic and method of science"². Thus, it broadly adopts the same philosophical stances as the natural sciences in seeking to explain and understand psychological phenomena. Unobservable psychological phenomena are understood as having an objective existence, and are amenable to study by empirical and experimental methods to understand their relations to the observable world.

3.2 RESEARCH DESIGN

² <http://www.psychologicalscience.org/index.php/publications/observer/2013/may-june-13/the-either-or-of-psychological-science-a-reflection.html>

Methods used to study the disposition effect were reviewed in the previous chapter. However, they will be revisited here, to explain how the methods used in this thesis link with the research philosophy described above.

As discussed above, an interpretivist approach has not been adopted. This thesis does not attempt to gain a qualitative understanding of how investors interpret or relate to the disposition effect. Qualitative approaches such as interviews, case studies, ethnography etc. are common when an interpretivist approach is adopted, but not as commonly used without it. Most research on the disposition effect is quantitative, and uses one of two methodologies: analysis of secondary data, or experiments. Both methodologies tend to follow the scientific method: forming hypotheses with deductive reasoning, then testing them by operationalising the variables in hypotheses. This thesis adopts the same approach.

Secondary data usually involves analysing trading data from real world financial markets. Its strength is that it directly measures trading behaviour in an environment where it occurs. If the research question is simply to prove that a disposition effect occurs in a specific market, then this can be demonstrated easily by interpretation of the observed behaviour. However, one weakness of this approach is when research is interested in the association of the disposition effect with other independent variables. These variables often must be estimated with proxies, weakening the external validity of the conclusions of the study.

A greater weakness of secondary data though, for this thesis, is that the researcher cannot directly manipulate variables. Even if independent variables are measured with 100% accuracy, secondary data can only show association rather than causation. So, using secondary data lacks internal validity in the conclusions drawn. Richards (2012) used a mixed method of secondary data and questionnaires to show that investors who habitually use reappraisal have some reduction in disposition effect. The use of questionnaires is an improvement on solely secondary data; however, it still only shows

association between variables. There could be some other reason why investors who habitually use reappraisal also show a difference in disposition effect.

The third and fourth research questions test the effect of cognitive reappraisal on the disposition effect. To improve the internal validity of this test an experimental design is adopted. An experimental group is compared with a control group to assess the effect of reappraisal. Pre-intervention and post-intervention measurements of trading behaviour are used to measure any change in behaviour. Using this design, there is strong internal validity in attributing causation of changes in the disposition effect to reappraisal.

This design also has an additional benefit, related to the motivation for researching reappraisal's effect on the disposition effect. The main motivation is theoretical, to investigate whether and how reappraisal changes the disposition effect. However, a secondary motivation is to show that reappraisal is a practical method that investors can use to improve their decision making. An excellent way to demonstrate this is by asking investors with no experience of reappraisal to begin implementing it.

The compromise the increased internal validity from the experimental design is decreased external validity. Trading behaviour is measured using a realistic stock market simulation game (the two-index game), rather than measuring it from trading decisions. However, to mitigate this, this thesis makes additional improvements in the external validity of the experiments carried out. In addition, the Milan study used to support the third research question uses retail investors as participant, rather than students or laymen. This increase in external validity while maintain internal validity from an experimental design is one of main the contributions the thesis aims to make, as explained in the previous chapter.

The first and second research questions also use the experimental data, but by combining it with data from the scale and trading data. The first research question aims to demonstrate that the disposition effect is a stable bias which shows convergent validity. The aim to show the existence of the disposition effect as a bias is motivated by

the realist perspective adopted: to show that the disposition effect is not merely a pattern of behaving, but is driven by an underlying trading bias.

To demonstrate intra-individual stability, the trading biases are measured multiple times in the research design. However, to demonstrate convergent validity, two methods of measuring trading biases are needed, so that variables driven by the same proposed bias can be compared. So, in addition to the two-index game, participants also completed the disposition effect scale. This asks about their behaviour and attitudes when trading in financial markets. Further discussion of this method is included in chapters 3 and 4.

The second research question aims to demonstrate that cutting gains and holding losses are stable biases, which show convergent and discriminant validity. Again, the aim to show the existence of these biases is motivated by the realist perspective adopted. The method used is an extension of that used for the first research question, in that it combines data from the two-index game and the disposition effect scale. In addition, factor analysis is performed with the scale to produce separate variables for cutting gains and holding losses.

This research design is based on the multitrait-multimethod matrix (MTMM), developed by Campbell and Fiske (1959). Intra-individual stability, convergent validity and discriminant validity are demonstrated by producing evidence for three assertions. Variables produce stable scores when repeatedly measured. Constructs which are theoretically related to each other (by latent traits) are observed as related to each other. Constructs which are not theoretically related to each other (because they measure different latent traits) are not observed as being related to each other. This is demonstrated by assessing the pattern of correlations between the same and different constructs, and comparing these patterns to those expected from theory.

In summary, two research designs are adopted in this thesis. Both are empirical and test research questions developed in the previous chapter. The first and second research

questions combine experimental data with a scale and trading data, and are based on the multitrait-multimethod matrix (MTMM). The third and fourth research questions use an experimental design that combines high internal validity when testing reappraisal, with increased external validity of that test.

3.3 METHODS

As discussed in the previous chapter, a main contribution of this thesis is to test cognitive reappraisal with greater external validity. This section describes how this is achieved. External validity is increased in two ways. Firstly, by increasing the ecological validity of how trading behaviour is measured in the lab. Second, by testing cognitive reappraisal using more representative participants. This section discusses how these improvements are made.

The main focus is the two-index game: this is the main method of measuring trading behaviours in the thesis and is used extensively in all three studies. This provides a key improvement compared with previous research on the disposition effect. The game is described, and it is explained why using it increases the ecological validity of the studies. The measurement of the disposition effect, cutting gains and holding losses is explained using it is also discussed.

The second part of improving the external validity of the studies is using more representative samples of participants in experimental work. This thesis includes two samples of retail investors, and one working adult sample. Further details of participant samples, recruitment and incentives are provided.

The experimental protocols are similar between all three studies. These are sketched out and contrasted, though more detail is given in later chapters. Participant incentives across the studies are also discussed.

3.3.1 The Two-Index Game

In the previous chapter, concerns were raised about the ecological validity of previous research on the disposition effect, which uses unrealistic environments for decision making. To build and improve on previous research, this thesis uses a stock market trading simulation called the two-index game ("the TIG"). The two-index game was developed as part of the xDelia project, an EU FP7 project which also studied the disposition effect, and adapted to suit the purposes of this thesis. This trading simulation significantly increases the ecological validity of the decision environment where trading decisions are made by participants. The nature of the game and how it achieves this is discussed further in this section.

The TIG presents participants with two moving indices (thus the name). One index shows the current price of a stock, and the other gives predictive information about future price movements. These are called the value index and predictor index, respectively. Figure 3.1 below is a screenshot of the game during play.

Figure 3.1 Screenshot of the two-index game



One drawback of previous experimental work is that participants had to respond to fixed prices: the price at the point they made a selling decision did not change while they were deciding what to do. Even in the case of Weber and Camerer (1998), who had 14 decision points, they were still fixed. This meant that participants do not have to deal with continuously varying prices. The two-index game improves on this by continuously changing the price of the stock being traded in real-time: the value index changes continuously as it would in real financial markets.

A related drawback of previous experimental research is that participants had to make decisions about holding or selling positions at fixed points. This experimental design is easier to administer, since there is no need to allow participants to make trades at any point. It is easy to setup a comparison of decisions about gains and losses, when choices are limited and the researcher can control whether a participant faces a gain or loss. It also simplifies data analysis, if the options participants consider are known in advance.

However, this approach reduces ecological validity, compared with trading in financial markets. Investors can trade markets in real time and can make decisions at any point. The two-index game attempts to capture this flexibility by also allowing participants to buy and sell positions at any time during the game. This allows them to consider buying, holding and selling shares in response to a continuously changing price. This is intended to be more cognitively demanding and emotionally arousing, because there is never a point in the game where participants can stop attending to the game: every moment of the game can be spent thinking about whether to hold or sell shares (or to buy shares if they are not currently being held).

Other features which increase ecological validity are the ability to short-sell and to take positions of varying sizes. Positions can be bought in sizes of 1, 3, 5 or 10 units (though some other studies do allow positions of different quantities to be taken). Short-selling allows investors to make money when the current price falls, by short-selling stocks at one price then closing that position at a lower price. This is a mirror image to how

investors make money by buying stocks at one price then closing that position (i.e. selling them) at a higher price. Short-selling has not been incorporated into previous experimental papers on the disposition effect. Although short-selling is not analysed separately from conventional trading, the fact that participants could trade this way, and these decisions were included when calculating a disposition effect, is itself a novel feature of the studies in this thesis.

3.3.1.1 The predictor index

In addition to a lack of available options when making trading decisions, another significant drawback of previous experimental work is the information participants are given when deciding whether to hold or sell stocks. This has been much simpler than in financial markets, reducing the ecological validity of those studies. The cognitive demands on participants have been much lower than investors in financial markets, so decisions in experiments are not made in the same environment that investors face.

Often participants have only price information to use in making decisions. This arguably encourages participants to display a disposition effect. Participants know that an experiment of this kind involves them making decisions using the information available. However, when little information is provided by the researchers, it implies that they should use the information they have available, which includes whether they have made a gain or loss. So, the less information that is provided, the more the setup encourages them to act on information about gains and loss, generating a disposition effect. A disposition effect may be produced by participants, but it may lack external validity (i.e. would not be replicated in real trading) since it is produced in a situation that lacks ecological validity.

Even studies which provided some form of predictive information have simple setups compared with financial markets. For example Lee *et al.* (1998) used 'analyst forecasts' which indicated that a stock was likely to rise (or fall) in the next period. These were presented as fixed predictions at each decision point. In Weber and Camerer (1998)

participants had a selection of stocks to trade, and were told that stocks had pre-existing probabilities of rising or falling in each trading period.³ This is certainly a more complicated setup, but this information still did not change in real-time: the probabilities were fixed at the beginning of the experiment and were unchanged over the 17 discrete trading periods participants observed. Since the probabilities were known and fixed, it is also a scenario about dealing with risk, rather than uncertainty, using the definitions from Knight (1921). Financial markets involve dealing with uncertainty rather than risk, and so does the two-index game.

The TIG improves on this by presenting participants with a prediction index. This has two features that greatly increase the cognitive demands of using this information. First, it changes in real-time like the price index does. Second, the relationship between the two indices is not revealed to participants, so there is high uncertainty about how to use the information in the predictor index to make decisions. This predictor index conceptually represents the many sources of information (in addition to the price history) which investors use when they make trading decisions. These sources of information also change continuously and have high levels of uncertainty in their implications for stock prices.

Movements in the value index are generated with a combination of random processes and information from the predictor index described below (though this was not revealed to participants) and are designed to mimic the movements found in financial markets. Participants were told (truthfully) that the predictor index provided some information about future movements of the price index, but that it was up to them to decide how to use it. In practical terms, using the predictor index was extremely difficult, increasing the complexity of the task given to participants. So, the predictor index makes the game

³ Participants were not told which probabilities related to which stock. This was left for the participants to estimate, by analysing the price movements of each stock over successive fixed trading periods.

more cognitively demanding; it also increases its similarity to a real trading environment, both of which improve ecological validity.

The basic relationship between the indices is that the value index copies previous movements of the predictor index with some delay, so that the predictor index moves a little ahead in time of the value index. However, there are several complications added to this trend to mask this relationship and make it more cognitively demanding to use.⁴ The time delay between indices changes in duration using a sinusoidal function, so that participants cannot anticipate future price movements simply by looking a fixed time ahead in the predictor index.

The value index does not always mirror changes in the predictor index. The game engine uses 10 "ticks" per second. On each of these ticks, the predictor index may rise, fall, or stay the same. Whether the value index copies the predictor index movement on each tick is based on a binomial distribution (to copy or not copy). This produces unreliability in the predictive signal that the predictor index generates.

The movement in the predictor index is also multiplied by a constant when that the direction is copied but not the same magnitude. Periodically, the value index moves in the opposite direction to the predictor index. In addition to all these, random noise (Brownian motion, Merton (1971)) is added to both indices.

Overall, the relationship between the two indices is difficult to follow with any accuracy. The main purpose of the predictor index was to make participants think about incorporating information other than the purchase price into their decisions in real time. This much more closely represents how decisions are made on stock markets.

⁴ Full details of how the indices behave are given in Lins & Yee (2011) and Yee & Lins (2011).

3.3.2 Measurement of the disposition effect

Playing the game produces a record of all the trades made during the game. Using this 'trading record' a disposition effect can be calculated for each participant for each play of the game. In effect, it allows a sample of each participant's trading patterns to be measured in a short space of time in an experimental setup. This allows experimental interventions to be tested experimentally, whilst still maintaining a high degree of ecological validity.

To demonstrate a disposition effect, it is necessary to show a difference in the selling of gains versus losses. There are two main methods of doing so in the literature. Odean's (1998) PGR-PLR method is commonly used in both field and experimental studies, and compares the relative frequency of selling in gains and losses. The main alternative to Odean's method is survival analysis. These methods are sketched out, before discussing how Odean's method is adapted for use in this study.

3.3.2.1 *Survival analysis*

A disposition effect, meaning cutting gains and holding losses, will result in losses being held longer than gains, since they are less likely to be sold at any given time. Survival analysis is a regression-based method used by Feng and Seasholes (2005). It models how long a stock will be held for; Feng and Seasholes use the number of days a stock is held from its initial purchase until it is sold entirely.

Parameters in the model adjust the probability that a stock will be held on any given day. Dummy variables code whether a stock is being held at a gain or loss, compared with the purchase price. The trading gains indicator (TGI) takes a value of 1 on any day where a stock is trading at a paper gain, or is sold at a gain. Conversely the trading loss indicator (TLI) takes a value of 1 on any day that a stock is trading at a paper loss, or is sold a loss. These are compared to the baseline, when a stock is deemed to not be trading at a gain or loss (when the purchase price falls between the daily high and low market prices). The disposition effect is demonstrated when these parameters

significantly affect the probability that a stock is held on any given day, in the direction expected with a disposition effect. So TGI decreases the probability a gain is held (i.e. it is more likely to be sold) and TLI increases the probability a loss is held (i.e. it is less likely to be sold).

This method is useful when the focus of a study is either controlling for or testing the influence of individual continuous variables, which are more easily dealt with in a regression model; however, this is not the case here.

3.3.2.2 PGR-PLR

Odean's PGR-PLR method compares the proportion of gains realised ("realised" meaning sold) with the proportion of losses realised. These are abbreviated to PGR and PLR. Each proportion is calculated as the number of observed sales at a gain or loss, as a proportion of the potential sales at a gain or loss. So, PGR and PLR represent the frequency of selling gains and losses, after adjusting for the opportunity to make those sales given the stocks in an investor's portfolio.

A stock which is sold is compared with the purchase price and classified as a realised gain or loss. For other stocks in the portfolio, their purchase price is compared to the daily market high or low price. A stock purchased for a price below above the daily market low is classified as a paper gain. A stock purchased for a price above the daily market high is classified as a paper loss. Using market high and low prices is a simplification to avoid having to use real time price data rather than daily data for stock prices.

PGR and PLR are defined as:

$$PGR = \frac{\text{Number of realised gains}}{\text{Number of realised gains} + \text{number of paper gains}}$$

$$PLR = \frac{\text{Number of realised losses}}{\text{Number of realised losses} + \text{number of paper losses}}$$

The disposition effect implies that losses will be held more than gains. It follows that if losses are held longer, they must be sold less frequently on average. However, an investor with many more losses in their portfolio than gains would be more likely to sell any one loss than any one gain. Using information about potential gains and losses takes this into account. A disposition effect is demonstrated when the proportion of available losses sold is lower than the proportion of available gains sold. This occurs when a loss is likely to be sold then again, *ceteris paribus*.

The disposition effect can be reported as a difference (PGR minus PLR) or a ratio (PGR divided by PLR). If a difference, PGR should be greater than PLR. A ratio shows the relative likelihood in selling a gain versus a ratio. A ratio of 1 means that gains and losses are equally likely to be sold. A ratio above 1 demonstrates that gains are more likely to be sold than losses. A ratio below 1 shows a "reverse disposition effect" where losses are more likely to be sold than gains.

In field studies, these ratios are usually calculated on each day during a trading record that stocks are sold (since on days that no trades take place, both proportions will be zero). Individual investors' disposition effects can be calculated by taking the average of each ratio over the whole period of a trading record. Alternatively, ratios can be aggregated over different periods of time, different classes of investors etc., as the research questions for each study require. Experimental studies usually make this comparison between groups. Participants are split into groups trading gains or losses, and disposition effect is demonstrated by showing a difference between those groups.

3.3.2.3 PGR-PLR adapted in this thesis

Odean's PGR-PLR terminology is used in this thesis for measurements calculated with the two-index game. However, PGR and PLR need to be adapted to be applied to the two-index game. The game produces a "trading record" analogous to the records used in field studies, but only one stock is traded. So, gains or losses sold as a proportion of the number of stocks available in a portfolio does not make sense.

As discussed earlier, the conceptual basis of PGR and PLR is to measure how frequently gains and losses are sold, after adjusting for the opportunity investors have to sell them. In the two-index game, the opportunity to sell them is measured by the total holding time for gains / losses, rather than the number of paper gains / losses. The holding time is measured in "ticks" during the game, which represent tenths of a second.

So, PGR and PLR are defined in this study as:

$$PGR = \frac{\text{Number of realised gains}}{\text{Total holding time of gains (in ticks)}}$$

$$PLR = \frac{\text{Number of realised losses}}{\text{Total holding time of losses (in ticks)}}$$

The disposition effect is defined as a ratio of these variables:

$$DE = \frac{PGR}{PLR}$$

This measurement of holding time is possible because of the flexible options that participants have when playing the game. Unlike other experimental studies, participants could hold positions for as long as they wished, and the holding period can be precisely measured and compared to other participants. One criticism that Feng and Seasholes (2005) makes of Odean's method is that Odean neglects all information from days where trades do not take place. They use survival analysis to allow this information to be used. However, the two-index game also allows all the information available in the trading record to be used.

In this thesis, DE is used to refer to a disposition effect score which has been calculated using data from the two-index game. Similarly, PGR and PLR refer to scores from the two-index game. The ratio is used since it has a more intuitive interpretation than the absolute difference between PGR and PLR. This is especially so in this study: the use of

holding time in the denominator makes the absolute figures for PGR and PLR very low and not intuitively meaningful.

PGR and PLR tend to be small, since the holding time figure is much greater than the number of trades made in each game. In addition, they are both bounded by zero, so their distributions have positive skew; to produce approximately normal distributions of PGR and PLR each variable is log transformed. DE is the ratio of the variables; however, DE is also log transformed. The terms PGR, PLR and DE refer to the log transformed variables unless noted otherwise.

An investor has no disposition effect where PGR is equal to PLR. Before logging, this produced a ratio 1. Since it is a ratio, DE greater than 1 corresponds with a disposition effect, and represents how many times more likely it is that a gain is sold compared with a loss (after adjusting for the opportunity to realize each type). DE is equal to PGR/PLR , so after logging, $\log(DE)$ is equal to $\log(PGR/PLR)$, which is equal to $\log(PGR)$ minus $\log(PLR)$. This means that an investor with no disposition effect has a DE of zero.

3.3.3 Participants

As discussed in the previous chapter, most experimental studies use students as participants. Experimental studies of trading decisions implicitly assume that their samples are representative of populations who trade on financial markets. To claim external validity, one sample must make trading decisions in the same way as the other. Studies involving emotion regulation make a similar assumption: the effect of emotion regulation on the sample and target population is the same. A contribution of this thesis is to increase the external validity of its conclusions by carrying out experiments with more representative participants. To achieve this, three samples of participants were used in the three studies included.

The London and Milan studies use retail investors. Retail investors actively trade their own capital, and are drawn from the same population studied in field studies of retail

investors. They have familiarity with the trading tasks used in these studies compared with the novice participant samples usually used in experimental studies.

A further aim of the thesis is to create interventions which can be used by retail investors to improve their financial decisions. So, these participant samples are drawn from the very same population of people that we wish to generalise the results to. In addition, of course, retail investors are more like the wider population of investors than students are.

The Milan study recruited participants at the trading conference "Trading Online Expo" run by Borsa Italiana (the Italian stock exchange) in central Milan. This Expo is aimed at private retail investors who trade on the stock market with their own capital. The Expo featured two days of parallel sessions of talks and seminars about trading and finance, and promotional stands around the main area mostly hosted by trading services companies. The Open University Business School also had a promotional stand, where participants were recruited and played the game for the first time.

Participants for the London study were recruited from two similar trade fairs in London: the London Investor Show in Olympia, and the World Money Show in the QEII conference centre. Both events in London are marketed as educational conferences aimed at investors trading privately with their own capital. As with Trading Online Expo there are also commercial stands, and a stand was used to recruit participants for the study. The stand was only used to explain the study to attendees and recruit participants, since the study was conducted entirely online. These participants are only used for the first and second research questions.

The OU study (at the Open University using a non-trader convenience sample) used a sample of the adults rather than retail investors. However, the sample is a moderate improvement on the common sample set of undergraduate students, having a broader demographic makeup. Participants came from doctoral students, academic staff and

non-academic staff from across the Open University campus. This also resulted in a broad range of ages.

3.3.4 Protocols

In the Milan study, participants initially filled out a short demographics questionnaire and gave their consent to take part (all instructions were doubled in English and Italian). Participants took an interactive video tutorial about how to play the game, and then played the game the first time (plat 1). The aim was that by playing the game once in person, it would discourage dropout later. After the Expo, participants were contacted by email to complete the study online. Between play 2 and play 3 participants answered the disposition effect scale, which is analysed in chapter 5. Further details are provided about the Milan study in chapter 6.

The London study protocol was very similar to the Milan one, except that no participation took place at the venues in London. They provided their contact details and were contacted online following the trade fairs, when they were prompted to complete the entire study.

The OU study was carried out in person on the Open University campus in Milton Keynes. Questionnaires were completed on paper, while temporary "labs" were set up using laptops around the campus to play the game.

Carrying out the study in person with layman was intended to increase sample size, improve data quality and reduce (or eliminate) dropout during the study. Compared with the other studies this allowed a relatively large sample to be collected, to more robustly replicate tests from the Milan study. The study also contained repeated questionnaires about emotions experienced. These may have been off-putting for retail investors, but were easily administered with the researcher in person. Further details are provided about the OU study in chapter 6.

A common feature of all 3 studies is that participants played the simulation four times in total, with each play generating a record of their trading decisions. Each game lasted 10 minutes in the Milan and London studies, and 5 minutes in the OU study. Multiple plays of the game had several aims.

The first aim was to allow multiple measurements of trading biases, for use in the first and second research questions. The second aim was to allow pre-intervention and post-intervention measurements, to assess the impact of cognitive reappraisal in the third and fourth research questions. The third aim was to give participants time to familiarise themselves with the mechanics of the game before they attempted to use cognitive reappraisal. This was particularly important for the lay sample in the OU study, who were obviously much less familiar with the process of trading than the experienced investor samples.

After completion, participants were emailed feedback on their performance and comparison to the group overall. They were also debriefed on the disposition effect and given feedback about their disposition effect too.

3.3.5 Participant Incentives

3.3.5.1 Milan and London studies

Monetary incentives were considered for the Milan and London studies; However, they were rejected for several reasons, both theoretical and practical. It was felt that small monetary amounts would not be sufficiently large to motivate investors who regularly traded much larger amounts of their own capital.

Providing small incentives may frame participation as “work” for which they were being lightly compensated, and it wouldn’t be worth their time to take part. If the purpose of taking part was perceived as the chance to win small amounts of money, participants may not engage seriously with the game if they feel the amounts at stake are not worth worrying about. Finally, providing only small incentives to take part might make the

study appear amateur, compared with the much larger amounts being discussed at the trading fairs and the professional firms on other promotional stands.

Instead of monetary incentives, participants were offered feedback on their performance in the game, comparing them with other participants. At recruitment, they were told the research was about the impact of emotions on their trading (without the detail about the disposition effect to avoid influencing outcomes participants). In their feedback, they were also de-briefed about true purpose of the study being to study the disposition effect, and its potential effect on their trading.

The studies were promoted as an opportunity to get insight into their own trading patterns and emotions during trading. This is often of great interest to retail investors; indeed, many were surprised that we weren't charging anything for the service! The prospect of feedback on their performance aimed to encourage participants to take the game seriously, despite their being no monetary rewards at stake. By explicitly comparing participants with each other, participants were encouraged to engage competitively, focus their attention on the game and treat it seriously.

3.3.5.2 OU study

The OU study was run more conventionally, providing incentives to take part for about an hour on campus. Participants were informed that a virtual lottery would take place following the study, and that a £40 voucher was available for every 5 people who took part in the study. Tickets to this 'lottery' were allocated based on scores for the third and fourth plays of the game. 120 participants were recruited; the highest scoring participant received 120 virtual "tickets", the second highest score received 119 tickets, and so on until the lowest scoring participant was given 1 ticket.

This incentive structure was chosen to fulfil multiple aims. Participants were encouraged to take the game seriously and try as hard as they could to get a good score. Offering a reward achieved this, while simply offering feedback on their trading may not have been

as appealing for a lay sample. Running a lottery provided an emotionally salient reward of £40, while keeping the average reward per participant affordable (£8).

Each play lasted for 5 minutes in the OU study. However, it was desirable to keep participants motivated through the game. Using a relative ranking to run the lottery gave participants this incentive. Even if they were performing poorly halfway through the game (or perceived themselves as performing poorly), performing well in the second half could still have a significant effect on their chances of winning by raising their relative rank.

Finally, a lottery gave every participant at least some chance of winning, even if they performed poorly. Without this, people who did not think they could win (for example if only a few prizes were given for the top scores) may have declined to take part.

3.4 CONCLUSION

This chapter has discussed the research philosophy, research design and some methods used in this thesis. This chapter began by discussing the research philosophy adopted. Very broadly, research can adopt subjectivism or some form of empiricism. This thesis contributes to the literature in behavioural economics and psychological science, and adopts an empiricist stance. It is interested in objectively studying trading behaviour, rather than studying how people construct their own understanding of trading behaviour. It applies the scientific method to carry out experiments, test hypotheses and draw conclusions.

Within an empirical research stance, this thesis adopts realism rather than positivism. The research here is interested in making statements about unobservable phenomena, by inferring knowledge about them from what can be observed.

Two methods are adopted. To address the first and second research questions, a combination of experimental data and questionnaire data is used. This is based on the multitrait-multimethod matrix developed by Campbell and Fiske (1959). Testing of the

cognitive reappraisal in the third and fourth research questions is carried with a classic intervention versus control experiment, using pre-intervention and post-intervention measurements.

The two-index game the main instrument used to measure the disposition effect. The mechanisms of the game were explained, and how Odean's PGR-PLR method is adapted to measure the disposition effect in the two-index game. Finally, details about the participants, protocols and participant incentives in the three studies included in this thesis were discussed.

4 TESTING THE NATURE OF TRADING BIASES USING THE TWO-INDEX GAME

This chapter discusses empirical evidence which concerns the first two research questions. Tests in this chapter use data from the two-index game to argue that the disposition effect, cutting gains, and holding losses all have trait-like characteristics. It does so by establishing: intra-individual stability of all three biases, convergent validity of the disposition effect, and discriminant validity between cutting gains and holding losses.

I begin by establishing that all three biases can be reliably measured, by analysing repeated measurements of all three biases using trading records from the two-index game. Intra-individual stability is assessed in three ways: by assessing patterns of bivariate correlations; by showing intra-individual stability in measurement over an extended period; and by estimating how much variance in a dataset of repeated measurements is attributable to differences between people. These tests are repeated for 3 independent samples (from the London, Milan and OU studies), producing strong evidence for intra-individual stability of these biases.

The chapter goes on to test the convergent validity of the disposition effect. Convergent validity is established where variables which are theoretically related to each other are in fact observed as being related to each other. This is another important piece of evidence in establishing that the disposition effect is an underlying individual bias, which exists independent of the situation it is measured, and influences behaviour across time and situations.

If an underlying bias does drive patterns of behaviour, then it should be possible to measure it consistently using different methods. This is demonstrated here by correlating scores from the two-index game with disposition effects calculated from retail investors' trading records in financial markets.

The chapter concludes by demonstrating that PGR and PLR have discriminant validity. Discriminant validity is the property that measurements of variables which are not theoretically related to each other are observed as not related to each other. This is crucial evidence for the hypothesis that cutting gains and holding losses are independent biases that have trait-like characteristics.

The combination of these two subcomponents, cutting gains and holding losses, defines an individual's disposition effect. However, it is argued not only that these biases have intra-individual stability, but that they are independent (or largely independent) biases. If they are independent, two individuals presenting the same disposition effect may display different levels of cutting gains and holding losses, even though they have the same overall disposition effect as other individuals.

Discriminant validity is assessed by the pattern of correlations between PGR and PLR from the two-index, across repeated plays. The pattern is compared to what would be expected if the biases were both associated with a unitary underlying disposition effect, and what would be expected if they measured different tendencies, and what would be expected if they were completely independent. An extension of this method controls for the overall frequency of trading as a confounding factor.

4.1 INTRA-INDIVIDUAL STABILITY OF TRADING BIASES IN THE TWO-INDEX GAME

4.1.1 Intra-individual stability of the disposition effect

4.1.1.1 *Intra-individual stability: Correlations between successive plays of the game*

If the disposition effect is a stable bias, it should be reliable over time. The disposition effect is measured here by the two-index game, whose output is processed to produce DE scores. Tables 3.1-3 show correlations for DE scores between successive plays are shown below from the London, Milan, and OU studies⁵. There are 4 plays from the

⁵ Sample sizes were 50, 46 and 103 respectively. Listwise deletion was used.

London and OU studies, producing 6 bivariate correlations in each, while there are 3 plays from the Milan study⁶ producing 3 bivariate correlations.

Table 4.1 London study - correlations between DE on successive plays

		Play 1	Play 2	Play 3	Play 4
Play 1	Correlation	-			
	p-value				
Play 2	Correlation	.750	-		
	p-value	<.001***			
Play 3	Correlation	.669	.798	-	
	p-value	<.001***	<.001		
Play 4	Correlation	.680	.635	.761	-
	p-value	<.001***	<.001***	<.001***	

Table 4.2 Milan study – correlations between DE on successive plays

		Play 2	Play 3	Play 4
Play 2	Correlation	-		
	p-value			
Play 3	Correlation	.849	-	
	p-value	<.001***		
Play 4	Correlation	.788	.787	-
	p-value	<.001***	<.001**	

Table 4.3 OU study – correlations between DE on successive plays

		Play 1	Play 2	Play 3	Play 4
Play 1	Correlation	-			
	p-value				
Play 2	Correlation	.497	-		
	p-value	<.001***			
Play 3	Correlation	.534	.679	-	
	p-value	<.001***	<.001***		
Play 4	Correlation	.486	.668	.793	-
	p-value	<.001***	<.001***	<.001***	

The overall pattern is that DE scores are highly correlated over multiple plays, showing that DE scores do measure a stable pattern in the way people make trading decisions in the game. Cohen (1988) suggested correlation coefficient of 0.1 represented a small

⁶ For the Milan study only plays 2, 3, and 4 are included. There are two reasons for this. First, many participants lost data for play 1, because of problems with wi-fi connectivity at the trade fair. Including play 1 would greatly reduce the sample size for these tests using listwise deletion. Second, participants who did have data for play 1 played this at the trade fair in Milan, and had a delay of at least a week before continuing with play 2 online. Since the timing is qualitatively different from the latter three plays, correlations with them are tested as discussed in the following section.

effect, 0.3 for a medium effect, and 0.5 for a large effect. The correlations here average around 0.7, which might be called a “very large” effect. Any subsequent references to the size of effects relate to Cohen’s suggestions, unless otherwise noted.

Cohen’s suggested benchmarks are often cited as hard rules, yet they were only suggestions (much like Fisher’s 0.05 criterion!) which need to be interpreted considering the source of the data and the goal of the study. The purpose of these correlations is to show that repeated measurements are driven by a shared underlying bias. It is argued here that correlations around 0.7 are definitely sufficient to plausibly make a case for that.

In the OU study, correlations between play 1 and the other 3 plays are lower than between other pairs. Since play 1 is the first attempt at playing the game, participants may have been paid more attention to learning about how the game works, and less attention to the positions of their trades (as gains or losses). This would diminish the expression of any disposition effect in their DE scores, which could account for the lower correlations. However, this effect is not apparent with play 1 in the London study. Participants in this study were retail investors, so this suggests they found the game less difficult to understand initially, and did not require much practice to play it fluently and express their normal patterns of decision-making.

4.1.1.2 Intra-individual stability over an extended period

In the London and OU studies, all four plays took place during the same session. However, for the Milan study, play 1 was played at recruitment in Italy, while plays 2-4 were played online a week or more later. This time difference allows testing of the stability of their disposition effect over a longer period. While the data is only available for 15 participants due to data loss, all these correlations are still strong, shown in table 4.4. Note that even the correlation with play 4 is still large, despite the additional noise generated by the experimental intervention between groups. So, this provides further good support that disposition effect scores are reliable over extended periods of time.

Table 4.4 Correlations of DE between play 1 with other plays in the Milan study

	Play 2	Play 3	Play 4
Correlation	.732	.702	.568
p-value	.002**	.004**	.027*

4.1.1.3 Variance in DE accounted for by differences between participants

A complementary analysis of intra-individual stability looks at how variance over successive plays can be split into between-participants and within-participants variance. High between-participants variance means that most variation in scores is due to differences between participants, and low within-participants variance means that each individual participant's scores do not differ much over repeated measurements. If most variation in scores is due to differences between participants, we can conclude that differences between people are stable and we have evidence for a bias that reliably differs between people.

Intercept-only random models were used to assess this. In this type of model, the total variance in a sample around the overall mean is divided into between-participant variance, and residual variance. (The overall mean is known as the intercept, thus the name of the model). Each participant has an individual mean for their scores (the distribution of participants' individual means is the "random factor" in the model). The deviations of individual means from the overall mean of the sample represent the between-participant variance.

Residual variance is calculated from the differences between each participant's observed scores and their individual means. Residual variance means the same here as in any regression model; However, since no other predictors are included in the model except individual means (the random factor), the residual variance represents all other sources of variation in scores. For example, these sources include random variation in an individual's score each time it is measured, and learning effects from playing the game multiple times.

The between-participants variance, as a percentage of the total variance, approximates how much variance in scores can be accounted for by differences between participants. The higher this is, the more evidence we have that scores reliably differ between participants and those differences represent a continuous distribution in an underlying disposition effect bias.

Table 4.5 below summarizes the results from the 3 studies. The percentage of between-participants variance varies by sample but the overall pattern is clear: the majority of variance in scores is due to differences between participants. This provides good evidence that the disposition effect is a stable bias expressed during decision making.

Table 4.5 Estimates of between-participant variance in DE

	Between-participant variance	Residual variance	% variance between-participants
London	.292	.123	70.3%
Milan	.316	.099	76.1%
OU	.120	.078	60.6%

4.1.2 Intra-individual stability of cutting gains

4.1.2.1 Correlations between successive plays of the game

Like the disposition effect, if cutting gains is a stable bias it should be reliable when measured repeatedly. In the two-index game, cutting gains is measured by PGR. As with the disposition effect, high correlations are expected between repeated plays of the game. Tables 3.6-3.8 show correlations for PGR are shown below for the London, Milan, and OU studies⁷.

⁷ Sample sizes were 50, 47 and 107 respectively. Listwise deletion was used. As discussed earlier, for the Milan study only plays 2-4 are included, and correlations with play 1 are analysed separately.

Table 4.6 London study – correlations between PGR on successive plays

		Play 1	Play 2	Play 3	Play 4
Play 1	Correlation	-			
	p-value				
Play 2	Correlation	.524	-		
	p-value	<.001***			
Play 3	Correlation	.380	.572	-	
	p-value	.007**	<.001***		
Play 4	Correlation	.344	.570	.746	-
	p-value	.015*	<.001***	<.001***	

Table 4.7 Milan study – correlations between PGR on successive plays

		Play 2	Play 3	Play 4
Play 2	Correlation	-		
	p-value			
Play 3	Correlation	.870	-	
	p-value	<.001***		
Play 4	Correlation	.790	.790	-
	p-value	<.001***	<.001***	

Table 4.8 OU study – correlations between PGR on successive plays

		Play 1	Play 2	Play 3	Play 4
Play 1	Correlation	-			
	p-value				
Play 2	Correlation	.719	-		
	p-value	<.001***			
Play 3	Correlation	.623	.790	-	
	p-value	<.001***	<.001***		
Play 4	Correlation	.551	.709	.806	-
	p-value	<.001***	<.001***	<.001***	

PGR scores are highly correlated over multiple plays of the games, and all correlations are significant. In the London study, correlations are mostly large effects ($\sim .05$), rather than very large ones, and the correlations are weaker with play 1. However, the overall pattern is clear.

4.1.2.2 Intra-individual stability over an extended period

As discussed in the disposition effect results above, correlations of PGR between play 1 and the other plays in the Milan study have been tested separately, since they allow

testing of stability over a longer period. Table 4.9 shows these correlations. As with the disposition effect results, despite the small sample size of 15 all correlations are significant and very large, providing further evidence for the stability and intra-individual stability of cutting gains as a bias.

Table 4.9 Correlations of PGR between play 1 with other plays in the Milan study

	Play 2	Play 3	Play 4
Correlation	.745	.699	.605
p-value	.001**	.004**	.017*

4.1.2.3 Variance in PGR accounted for by differences between participants

The variance attribution method described earlier for DE scores is also used here for PGR scores, and the results are shown in table 4.10. The results confirm that most of the variance in PGR scores in the three studies is attributable to between-participant differences. This provides good evidence that cutting gains is a stable bias during decision making. Between-participant variance in the London study is lower than expected, consistent with the lower correlations using play 1; however, it still makes up the majority of variance observed in scores.

Table 4.10 Estimates of between-participant variance in PGR

	Between-participant variance	Residual variance	% variance between-participants
London	.053	.042	55.8%
Milan	.123	.032	79.4%
OU	.125	.60	67.6%

4.1.3 Intra-individual stability of holding losses

4.1.3.1 Correlations between successive plays of the game

Like the disposition effect and cutting gains, if holding losses is a stable bias then it should be reliable when measured repeatedly. In the two-index game, holding losses is measured by PLR. Again, high correlations are expected between repeated plays of the game. Tables 3.11-13 show correlations for PLR from the London, Milan, and OU

studies⁸. PLR is highly correlated over multiple plays of the game. Correlations are even higher than for DE and PGR scores, with most more than 0.7 and many approaching 0.9. Overall, scores for PLR are extremely reliable across these 3 independent samples.

Table 4.11 London study – correlations between PLR on successive plays

		Play 1	Play 2	Play 3	Play 4
Play 1	Correlation	-			
	p-value				
Play 2	Correlation	.785	-		
	p-value	<.001			
Play 3	Correlation	.712	.852	-	
	p-value	<.001	<.001		
Play 4	Correlation	.738	.709	.830	-
	p-value	<.001	<.001	<.001	

Table 4.12 Milan study – correlations between PLR on successive plays

		Play 2	Play 3	Play 4
Play 2	Correlation	-		
	p-value			
Play 3	Correlation	.891	-	
	p-value	<.001		
Play 4	Correlation	.843	.803	-
	p-value	<.001	<.001	

Table 4.13 OU study – correlations between PLR on successive plays

		Play 1	Play 2	Play 3	Play 4
Play 1	Correlation	-			
	p-value				
Play 2	Correlation	.661	-		
	p-value	<.001			
Play 3	Correlation	.668	.802	-	
	p-value	<.001	<.001		
Play 4	Correlation	.645	.801	.876	-
	p-value	<.001	<.001	<.001	

⁸ Sample sizes were 50, 47 and 104 respectively. Listwise deletion was used. As discussed earlier, for the Milan study only plays 2-4 are included, and correlations with play 1 are analysed separately.

4.1.3.2 *Intra-individual stability over an extended period*

As discussed, correlations of PLR between play 1 and the other plays in the Milan study have been tested separately, since they allow testing of stability over a longer period. These results are shown in table 4.14. Again, despite the reduced sample size (n=15), all correlations are large and significant, further supporting the stability of holding losses over time.

Table 4.14 Correlations of PLR between play 1 and other plays in the Milan study

	Play 2	Play 3	Play 4
r	.657	.597	.585
p	.008**	.019*	.022*

4.1.3.3 *Variance in PLR accounted for by differences between participants*

The variance attribution method described earlier is used here again for PLR scores, with the results shown in table 4.15. The high between-participant variance confirms that most of variation of PLR scores is also attributable to between-participant differences, again providing evidence that scores represent measurement of a stable bias.

Table 4.15 Variance in PLR scores accounted for by differences between participants

	Between-participant variance	Residual variance	% variance between-participants
London	.318	.094	77.2%
Milan	.329	.084	80.0%
OU	.281	.106	72.6%

4.2 CONVERGENT VALIDITY OF THE DISPOSITION EFFECT

The section above has established that the disposition effect can be reliably measured over time. This section tests whether retail investors display a similar propensity for the disposition effect when it is measured using different methods. As noted earlier, this

tests convergent validity of the disposition effect: if the disposition effect is an underlying bias of an individual, it should be possible to measure it in multiple ways and still find consistent results. For example, investors who score highly on one measure should tend to score highly on another measure.

Disposition effect scores from the two-index game are compared with disposition effect scores from the trading records of retail investors, recording 3 months of activity in financial markets. This data was supplied by Saxo bank, who have collaborated extensively with the Open University (and Prof. Fenton-O’Creevy), and who developed the two-index game. The data analysis was performed by Paul Grayson.

This correlation is significant, with a medium effect size ($r = .271$, $p = .026$, $n = 68$).

Although this result is brief, it is important in the context of the wider thesis. This result demonstrates convergent validity of the disposition effect in two different situations.

While the two-index game is designed to mimic trading, there are obviously differences between a short trading simulation and real-world trading over an extended period. The fact that scores from the two methods correlate is strong evidence that a disposition effect bias exists.

In addition, this result supports the use of the two-index game, which is the primary method used to measure trading biases in this thesis. Although this thesis carries out experimental studies, it is ultimately concerned with the disposition effect as observed in real-world trading. By showing that the two-index game correlates with disposition effects from real-world trading, its use as an experimental proxy of the bias is justified.

4.3 DISCRIMINANT VALIDITY ACROSS PLAYS OF THE TWO-INDEX GAME

Discriminant validity is the concept that variables which are not theoretically related to each other are also observed as not related to each other. This property relates to the hypothesis that cutting gains and holding losses are independent biases.

Although the evidence above shows that cutting gains and holding losses are stable biases, they may or may not *a priori* be independent of each another. The stereotyped trading pattern of someone with a disposition effect is someone who cuts gains and holds losses. This assumes that people will tend to show both behaviours, or express neither.

Anyone with a difference in the likelihood of selling gains and losses has a disposition effect (by definition). However, if cutting gains and holding losses are independent biases, their contribution to an individual's disposition effect may vary from person to person. For example, someone who sells losses slowly may trade gains similarly to other people. In this case they would display a bias for holding losses, but not for cutting gains. Conversely, they may sell losses similarly to other people but sell gains more quickly, so they would display a bias for cutting gains.

The two-index game produces data for both PGR and PLR on each play, so it is also possible to test discriminant validity between them. The hypothesis being tested is that the propensities for cutting gains and holding losses are independent, so if cutting gains and holding losses are measured by PGR and PLR, it follows that PGR and PLR should be independent of each other.

4.3.1 Testing discriminant validity with the two-index game

Discriminant validity is established by using multiple scores from the game for PGR and PLR, and comparing them with each other. Each participant generated 3 or 4 scores each of PGR and PLR during their participation in the study. These scores can all be correlated with each other. The strength of the correlations for different categories of correlations is evidence for or against discriminant validity.

Correlations are within-bias or between-bias. A within-bias correlation measures the correlation between two sets of scores for the same bias, for example PGR scores from play 2, and PGR scores from play 4. A between-bias correlation measures correlations between different biases, so between a PGR score and a PLR score. Within-bias

correlations have already been discussed above, when demonstrating intra-individual stability of PGR and PLR; however, they are used again now to compare with the between-bias correlations.

Discriminant validity is assessed by analysing the pattern of within-bias and between-bias correlations. If the biases have discriminant validity, i.e. measure different things, then the between-bias correlations will be low, and certainly lower than the within-bias correlations. So, the expected pattern is that within-bias correlations should be high, while between-bias correlations should be as low as possible. There are no strict cut-off points for the interpretation of these correlations matrices; evidence for discriminant validity is derived from the overall pattern which emerges from many correlations (Campbell and Fiske, 1959).

Between-bias correlations come in two further subcategories. All the within-bias correlations must be from different plays of the game, since PGR and PLR were only measured once per game. However, between-bias correlations can be from the same play, or different plays. For example, PGR from play 2 can be correlated with PLR from play 2, or with PLR from play 4. I refer to these as within-play and between-play correlations respectively.

Within-play between-bias correlations are derived from trading data from the same play. Each play may have specific factors that could affect PGR and PLR simultaneously, for example level of familiarity with the game (which increases over the study), how the player is performing (how much cumulative profit) during the game, etc. These are effectively shared noise when trying to measure the latent biases of cutting gains and holding losses. So, when PGR and PLR are measured at the same time, they may be expected to correlate more strongly than the between-play between-bias correlations, because of this "noise".

It follows that the strongest test of discriminant validity with this design is comparing within-play between-bias correlations (e.g. PGR₂ & PLR₂), with between-play within-bias

correlations (e.g. PGR2 & PGR4). The former category correlates different biases measured at the same time, while the latter correlate the same bias at different points in time. If within-bias correlations are stronger than the within-play correlations, this would provide strong evidence for cutting gains and holding losses being latent biases driving scores in PGR and PLR.

4.3.2 Analysis of correlation matrices from each study

Correlation matrices for each study are shown and discussed below. Correlation coefficients are in the first table for each study, and p-values in the second table⁹. To make the overall pattern easier to see, within-bias correlations have heavy shading, within-play between-bias correlations have light shading, and between-play between-bias correlations have no shading.

4.3.2.1 London study

There is a clear pattern that confirms the predictions made. Within-bias and between-bias correlations have little overlap in the strength of correlations or their significance levels. Correlations are much stronger for within-bias correlations, and all within-bias correlations are significant. All but two between-bias correlations are lower in strength than every single within-play correlation and 12 of 16 are non-significant.

The shared noise on each play of the game appears to have had an effect, since the within-play between-bias correlations are a little stronger overall than the between-play between-bias correlations. For example, the highest between-bias correlation for PGR1 is with PLR1; the highest between-bias correlation for PGR2 is with PLR2, etc. However, they are still clearly much lower in strength when compared to the within-bias correlations.

⁹ Sample sizes were 50, 47 and 103 respectively. Listwise deletion was used.

Table 4.16 London study - correlation coefficients between PGR and PLR in the two-index game

	PGR1	PGR2	PGR3	PGR4	PLR1	PLR2	PLR3	PLR4
PGR1								
PGR2	.524							
PGR3	.380	.572						
PGR4	.344	.570	.746					
PLR1	.289	.309	.092	.154				
PLR2	.259	.488	.115	.264	.785			
PLR3	.160	.401	.106	.263	.712	.852		
PLR4	.099	.274	.047	.159	.738	.709	.830	

Table 4.17 London study - correlation p-values between PGR and PLR in the two-index game

	PGR1	PGR2	PGR3	PGR4	PLR1	PLR2	PLR3	PLR4
PGR1								
PGR2	<.001							
PGR3	.007	<.001						
PGR4	.015	<.001	<.001					
PLR1	.042	.029	.526	.286				
PLR2	.069	.000	.425	.064	<.001			
PLR3	.268	.004	.464	.065	<.001	<.001		
PLR4	.495	.055	.744	.272	<.001	<.001	<.001	

4.3.2.2 Milan study

Again, there is a clear pattern that confirms what was predicted. Within-bias and between-bias correlations form two groups of correlations, and this time there is no overlap in either strength or significance values. Correlations are much stronger for within-bias correlations, and all within-bias correlations are highly significant. Every between-bias correlation is lower than every within-bias correlation.

A difference from the London data is that although between-bias correlations are much lower in comparison to the within-bias ones, most are significant. This is unexpected since if these measures were completely independent, correlations would be close to zero and not significant. The overall pattern still provides evidence that PGR and PLR have some discriminant validity, with clear differences in strength of correlation between

within-bias and between-bias correlations, but they are not completely independent in this dataset.

Table 4.18 Milan study - correlation coefficients between PGR and PLR in the two-index game

	PGR2	PGR3	PGR4	PLR2	PLR3	PLR4
PGR2						
PGR3	.870					
PGR4	.790	.790				
PLR2	.408	.344	.424			
PLR3	.317	.280	.326	.891		
PLR4	.271	.200	.358	.843	.805	

Table 4.19 Milan study - correlation p-values between PGR and PLR in the two-index game

	PGR2	PGR3	PGR4	PLR2	PLR3	PLR4
PGR2						
PGR3	<.001					
PGR4	<.001	<.001				
PLR2	.004	.018	.003			
PLR3	.030	.057	.025	<.001		
PLR4	.065	.177	.013	<.001	<.001	

4.3.2.3 OU study

The data from the OU study are more mixed. The first thing that stands out is that every correlation is significant. This is partly due to the increased sample size, but also because the between-bias correlations are higher than previously found. This is certainly unexpected, and means the evidence for discriminant validity between PGR and PLR is a lot weaker in this dataset.

Within-bias correlations are still a little higher than between- correlations. To give an idea of the overall pattern, we can look at the average correlation coefficient for each category of correlation. This is not a statistical test, but just gives an idea of what's going on. The average PGR-PGR correlation is 0.702, for PLR-PLR is 0.734 and for between-bias correlations (PGR-PLR) is 0.566. So, this still provides some evidence for discriminant validity, but not as strong evidence.

Splitting the between-bias correlations, the average within-play correlation is .673 (not much lower than the .702 average for PGR-PGR), while the average between-play correlation is .531. So as discussed above, there appears to be some shared variance for within-play correlations that is missing for between-bias correlations.

Table 4.20 OU study - correlation coefficients between PGR and PLR in the two-index game

	PGR1	PGR2	PGR3	PGR4	PLR1	PLR2	PLR3	PLR4
PGR1								
PGR2	.716							
PGR3	.625	.797						
PGR4	.556	.716	.806					
PLR1	.674	.494	.457	.389				
PLR2	.545	.687	.607	.515	.644			
PLR3	.513	.584	.689	.527	.659	.798		
PLR4	.506	.582	.648	.640	.635	.797	.873	

Table 4.21 OU study - correlation p-values between PGR and PLR in the two-index game

	PGR1	PGR2	PGR3	PGR4	PLR1	PLR2	PLR3	PLR4
PGR1								
PGR2	<.001							
PGR3	<.001	<.001						
PGR4	<.001	<.001	<.001					
PLR1	<.001	<.001	<.001	<.001				
PLR2	<.001	<.001	<.001	<.001	<.001			
PLR3	<.001	<.001	<.001	<.001	<.001	<.001		
PLR4	<.001	<.001	<.001	<.001	<.001	<.001	<.001	

4.3.2.4 Summary of results from all three studies

In the London study, within-bias correlations are clearly stronger than between-bias ones. The majority of between-bias correlations are weaker than every within-bias correlation, with a few exceptions. This is consistent with the latent biases driving these correlations (cutting gains and holding losses) having discriminant validity. There is some evidence of a within-play factor on correlations; However, within-play between-bias correlations are still lower than almost all within-bias ones, again supporting discriminant validity.

The Milan study delivers very similar results to the London study, except that correlations are stronger across the board. This results in most between-bias correlations being significant, though at lower levels than the within-bias correlations. This means the data is weaker evidence for independence of cutting gains and holding losses is weaker. However, the difference in correlation strength of within-bias and between-bias correlations is still strong evidence for discriminant validity.

The OU study produces results most different to the other two. Every correlation is highly significant, regardless of type. There is some evidence for discriminant validity between PGR and PLR, since the within-bias correlations are stronger than the between-bias correlations on average; However, there is considerable overlap in the strength of correlations between the two types, so the evidence for discriminant validity is much weaker than for the other two studies. There is also evidence for shared within-play variance, which is almost as strong as the between-bias shared variance.

So, there is strong evidence for discriminant validity between PGR and PLR in the London and Milan studies, but less evidence for independence of these biases, since there are still some significant correlations between PGR and PLR. In the OU dataset between-bias correlations are all significant, and almost as large as some within-bias ones, providing only weak evidence for discriminant validity.

4.4 DISCRIMINANT VALDITY CONTROLLING FOR TRADING FREQUENCY

Trading frequency is a potentially confounding factor when trying to assess discriminant validity using the technique above. PGR and PLR are both fractions: the number of sales made at a gain/loss, divided by time spent holding gains/losses. So, someone who trades a lot will tend to have higher PGR and PLR than average, because there will be many gains and losses sold compared with the population average. Conversely a person who trades less frequently will tend to have lower PGR and PLR than average, because they will sell few gains and losses compared with the population average.

This results in PGR and PLR being correlated partly because they are both affected by the overall frequency of trading. Some of the shared variance between them will be attributable to differences in how many trades someone tends to make, rather than their behaviour towards gains or losses specifically. This shared variance will mask the contribution of the latent biases we are interested in: cutting gains and holding losses.

To remove this confounding factor, PGR and PLR are regressed on the number of total trades made. This is carried out within each play; for example, PGR2 and PLR2 are both regressed on the total number of trades¹⁰ from play 2. The residuals of these regressions represent the variance in PGR and PLR, after the variance due to the frequency of trading has been removed.

Total number of trades is used for both PGR and PLR, rather than the number of gains or losses sold. Using the latter approach would remove information relating to differences between gains and losses; how frequently gains and losses specifically are sold is informative about differences in behaviour towards gains versus losses. Using the overall number of trades removes information about the frequency of trading, without changing information about the *relative* number of gains versus losses sold in each play.

4.4.1 Analysis of correlation matrices from each study

The correlation matrices presented in the previous section are now presented and discussed again, but using the regression residuals rather than the raw figures¹¹. Again, to make the overall pattern easier to see, within-bias correlations have heavy shading, within-play between-bias correlations have light shading, and between-play between-bias correlations have no shading.

¹⁰ As with PGR and PLR, TNT has been logged before analysis, in order to approximate a normal distribution.

¹¹ As before sample sizes were 50, 47 and 103 respectively. Listwise deletion was used.

4.4.1.1 London study

Controlling for trading frequency reproduces the pattern previously seen in the London study, and makes the evidence for discriminant validity much stronger. Most between-bias correlations are now close to zero, indicating that PGR and PLR residuals are close to independent. All but one of the between-bias correlations are not significant. So, after controlling for trading frequency, we can say there is little to no association between PGR and PLR, which is what would be expected if they measure independent biases.

In contrast, the within-bias correlations are still strong and significant. In particular, the PLR correlations are all very strong. This is good evidence that trading frequency was responsible for the between-bias correlations seen previously, but that the latent biases of cutting gains and holding losses continue to drive the within-bias correlations even after controlling for this.

Table 4.22 London study - correlation coefficients between PGR and PLR residuals in the two-index game of residuals after regression on trading frequency

	PGR1	PGR2	PGR3	PGR4	PLR1	PLR2	PLR3	PLR4
PGR1								
PGR2	.596							
PGR3	.395	.597						
PGR4	.323	.596	.770					
PLR1	.057	.219	-.008	.081				
PLR2	.218	.379	.115	.211	.786			
PLR3	-.011	.209	-.054	.133	.680	.709		
PLR4	-.085	.125	-.084	.020	.656	.578	.702	

Table 4.23 London study - correlation p-values between PGR and PLR residual in the two-index game after regression on trading frequency

	PGR1	PGR2	PGR3	PGR4	PLR1	PLR2	PLR3	PLR4
PGR1								
PGR2	<.001							
PGR3	.005	<.001						
PGR4	.023	<.001	<.001					
PLR1	.696	.130	.955	.582				
PLR2	.132	.007	.432	.145	<.001			
PLR3	.940	.150	.714	.361	<.001	<.001		
PLR4	.560	.391	.566	.892	<.001	<.001	<.001	

4.4.1.2 Milan study

As with the London study, this confirms the pattern previously seen in the Milan study, but makes the evidence stronger. All between-bias correlations are now close to zero, and none are significant, so the between-bias correlations can be treated as random variation. After controlling for trading frequency then, there is little to no association between PGR and PLR, as expected if they measured independent biases.

In contrast, within-bias correlations are all still very strong and highly significant. As with the London study, this is strong evidence that the latent biases of cutting gains and holding losses drive the within-bias correlations. Meanwhile, between-bias correlations were driven by trading frequency and not a shared disposition effect.

Table 4.24 Milan study - correlation coefficients between PGR and PLR residuals in the two-index game of residuals after regression on trading frequency

	PGR2	PGR3	PGR4	PLR2	PLR3	PLR4
PGR2						
PGR3	.865					
PGR4	.734	.751				
PLR2	.125	.101	.209			
PLR3	.136	.095	.183	.868		
PLR4	.049	-.038	.183	.812	.750	

Table 4.25 Milan study - correlation p-values between PGR and PLR residual in the two-index game after regression on trading frequency

	PGR2	PGR3	PGR4	PLR2	PLR3	PLR4
PGR2						
PGR3	<.001					
PGR4	<.001	<.001				
PLR2	.401	.498	.159			
PLR3	.361	.525	.217	<.001		
PLR4	.744	.800	.219	<.001	<.001	

4.4.1.3 OU study

The effect of controlling for trading frequency in the OU study is not as stark, but it does improve the evidence for discriminant validity. Between-bias correlations are not weakened as much as the previous two studies, where most between-bias correlations

became non-significant and close to zero. However, a clear pattern in the correlations emerges.

Within-bias correlations are all greater than 0.4, mostly greater than 0.6, and are highly significant ($p < .001$). Before controlling for trading frequency, between-bias correlations were of similar magnitude and all highly significant. After controlling for trading frequency, all coefficients are lower than 0.4 with an average coefficient of around 0.25, and every between-bias correlation is weaker than the weakest within-bias correlation. Most are still significant, so cutting gains and holding losses aren't demonstrated as completely independent of one another in this data. However, it is clear that within-bias correlations are consistently stronger than between-bias correlations, once we have controlled for trading frequency.

Table 4.26 OU study - correlation coefficients between PGR and PLR residuals in the two-index game of residuals after regression on trading frequency

	PGR1	PGR2	PGR3	PGR4	PLR1	PLR2	PLR3	PLR4
PGR1								
PGR2	.623							
PGR3	.564	.763						
PGR4	.503	.680	.820					
PLR1	.267	.208	.235	.222				
PLR2	.250	.358	.380	.351	.531			
PLR3	.173	.215	.272	.101	.445	.616		
PLR4	.187	.235	.334	.258	.475	.716	.723	

Table 4.27 OU study - correlation p-values between PGR and PLR residual in the two-index game after regression on trading frequency

	PGR1	PGR2	PGR3	PGR4	PLR1	PLR2	PLR3	PLR4
PGR1								
PGR2	<.001							
PGR3	<.001	<.001						
PGR4	<.001	<.001	<.001					
PLR1	.006	.035	.017	.024				
PLR2	.011	<.001	<.001	<.001	<.001			
PLR3	.081	.029	.006	.311	<.001	<.001		
PLR4	.059	.017	.001	.008	<.001	<.001	<.001	

4.4.1.4 Summary of results from all three studies

The overall pattern across the three studies is that controlling for trading frequency substantially improves the evidence for discriminant validity between PGR and PLR. In the London and Milan studies, most between-bias correlations are reduced to near zero and are not significant. Crucially, within-construct correlations are still high and are all highly significant. This difference is strong evidence for discriminant validity of PGR and PLR, and therefore the independence of cutting gains and holding losses.

The OU study is not as conclusive. Within-bias correlations are still strong and all highly significant, but most between-bias correlations are still significant too. However, there is a clear delineation of strength of correlations, with all between-bias correlations being below .4, while most within-bias correlations are above .6. So, this study also provides evidence for discriminant validity of PGR and PLR in showing that they measure different biases, though it provides weaker evidence for stating that PGR and PLR are completely independent of one another.

4.5 CONCLUSION

This chapter presented evidence relating to parts of the first two research questions. Using data from the two-index game it argued for the intra-individual stability of the disposition effect, cutting gains and holding losses, and found that all three biases were reliably measured. This was demonstrated using three different techniques: strong correlations over repeated measurements, stability over an extended period, and attributing variance to differences between people's biases. Similar results were found for all 3 independent studies, so this is a strong finding.

It is also explored the extent to which PGR and PLR show discriminant validity when tested with the two-index game. It found good evidence for discriminant validity in the London and Milan studies, though the evidence was not as strong in the OU study. After controlling for trading frequency, results from the London and Milan studies displayed

very clear discriminant validity, or even that PGR and PLR may be completely independent. This supports the hypothesis that cutting gains and holding losses are two independent biases which combine to produce a disposition effect. After controlling for trading frequency result from the OU study supported discriminant validity between PGR and PLR but did not support complete independence of the measurements.

5 ANALYSIS OF THE TRADING SCALE AND ITS RELATION TO SCORES FROM THE TWO-INDEX GAME

Like chapter 4 this chapter relates to the first two research questions, and continues to argue for trait-like characteristics of the disposition effect, cutting gains, and holding losses. It builds on the results from chapter 4 by introducing a scale used to measure trading behaviour. It uses data from the two-index game again, but combines this with a disposition effect scale, which was completed by retail investors in the Milan and London studies. This data is used to demonstrate convergent validity and discriminant validity of the disposition effect, cutting gains, and holding losses, which in turn supports viewing them as possessing trait-like characteristics.

The chapter begins by describing the scale and how disposition effect scores were created from it. Then these scores are correlated with DE scores from the two-index game. This is another demonstration of convergent validity of the disposition effect, which was also tested in chapter 4. If an underlying behavioural tendency is driving behaviour, then it should be possible to measure that tendency consistently using different methods.

The chapter goes on to analyse whether the scale can be split into two factors: one representing cutting gains and one representing holding losses. Inter-item correlations are explored, and possible groupings between them discussed. Then the scale is split into underlying factors using principal components analysis.

The ability to extract two factors demonstrates that the disposition effect has a dual nature, and that this is captured by the scale factors. In contrast, a factor solution that extracted only a single factor would suggest a monolithic disposition effect bias. (A factor solution producing no common factors would indicate the items do not measure any common behaviour and the scale is not a useful measurement of any latent tendency.)

After showing that two factors can be extracted, the factors are tested for independence. This provides evidence for discriminant validity between them: it is conceptually like the

tests in chapter 4, showing that cutting gains and holding losses from the two-index game do not correlate with each other. So, chapter 5 strengthens the evidence that cutting gains and holding losses are essentially separate biases, which in combination define a person's disposition effect.

Once the scale has been split into two factors, it is possible to test convergent and discriminant validity, using factor scores from the scale and two-index game scores simultaneously. This is done by correlating PGR and PLR from the two-index game with factor scores from the scale. As with the disposition effect, convergent validity can be established by testing whether the same bias measured with the two different methods produces correlated measurements. PGR is expected to correlate with the gain factor extracted, and PLR with the loss factor extracted.

Discriminant validity is tested in a similar way, but now opposing biases are correlated with each other. Since they are expected to measure different biases, correlations are **not** expected to be significant. PGR should fail to correlate with the loss factor extracted, and PLR should fail to correlate with the gain factor extracted. If successful, establishing convergent and discriminant validity can provide strong evidence for treating cutting gains and holding losses as independent biases in their own right, independent of the disposition effect.

5.1 DESCRIPTION AND PREPARATION OF SCALE DATA

London and Milan participants completed the scale as detailed in chapter 2. This is a Likert scale of 10 items that aims to measure common attitudes and behaviours which someone trading with a disposition effect may display. The items were picked for their face validity of the disposition effect. The scale was originally used during the xDelia project, which also researched the disposition effect in retail investors.

The scale is shown in table 5.1, and shows whether each item relates to gains or losses, and the type of trading pattern or attitude which it describes. The grouping of these items is discussed later in this chapter. Each item had a response scale from 1 to 5,

representing: "strongly disagree", "disagree", "neither agree, nor disagree", "agree", and "strongly agree" respectively.

Table 5.1 Items in the disposition effect scale

#	Item text	Gain/ Loss	Trading pattern or attitude
1	If a trade is succeeding I would rather close the trade than take risks for the chance of further gains.	Gain	Cutting gains
2	I am usually willing to take some extra risks to recover a loss.	Loss	Holding losses
3	I feel the effects of losses more than gains	Both	Loss aversion
4	When the value of a trade I have made falls, I am usually confident it will rise in value again	Loss	Holding losses
5	When the value of a trade I have made falls, I immediately close its position	Loss	Cutting losses (reversed item)
6	When the value of a trade I have made falls, I cut my losses without regret	Loss	Cutting losses (reversed item)
7	When the value of a trade I have made rises, I wait for it to drop before closing its position.	Gain	Holding gains (reversed item)
8	In trading, you need to take some risk when your trade is falling and not abandon it immediately.	Loss	Holding losses
9	When the value of a trade I have made rises, I always close its position before its value can fall again	Gain	Cutting gains
10	When the value of a trade I have closed rises still further, I accept the situation without regret.	Gain	Cutting gains

Briefly grouping the items, there are four items relating to gains, five relating to losses, and one explicitly asking about an investor's loss aversion, which is sometimes hypothesised to drive the disposition effect.

5.1.1.1 Data preparation

There is no clear answer to what constitutes sufficient numbers for carrying out principal components analysis or factor analysis. A common suggestion is that 10 respondents per item is a reasonable amount of data, though more is always desirable. A larger total sample size is always better, but fewer than 100 (regardless of the number of items) are often considered inadequate.

The scale data from the Milan and London studies are pooled together for analysis, giving a sample of 108 participants in total (1 participant completed the two-index game plays but did not complete the scale, so is omitted). Complete data is necessary to carry out principal components analysis. Listwise deletion of cases would reduce the sample to 98, resulting in approximately 10% attrition. However, the amount of data missing is relatively small: only 11 responses are missing from a total of 1,080, just over 1% of the total information.

To retain as many cases as possible, these missing responses have been imputed using multiple imputation. This is widely considered the new gold standard when dealing with missing data (Allison, 2002). Rather than impute a single estimate, as mean imputation or regression imputation do, multiple imputation creates multiple complete datasets, filling in missing data with new estimates in each dataset. The estimates differ as they are drawn from stochastic distributions of the estimated values, rather than the existing data in the dataset completely determining the imputed values (as with the mean and regression methods). So, creating multiple estimates allows complete datasets to be created, while maintaining the error that the data are estimated with. Current thinking is that 5 imputations can produce reasonable estimates. To be conservative, 10 complete datasets have been imputed.

5.1.1.1.1 Correlations after using multiple imputation

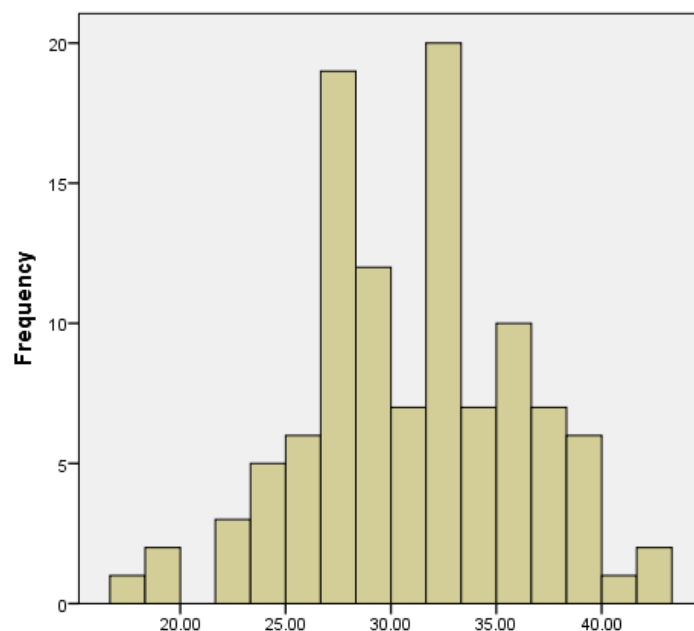
Multiple imputation solves the issue of missing data, but it creates the problem of how to use multiple datasets to analyse the scale. This is solved by pooling the results of all imputations. The best estimate of the true correlation coefficient is simply the average coefficient over all imputations. The process for p-values is more complex. The process begins with the average p-value, but is adjusted up, to avoid a p-value which is too low. The logic behind this is that since missing data are imputed multiple times, an outlier estimate in one dataset could produce a particularly low p-value for that imputation, which would not be reflective of the true relationship in the data. The mathematics of this adjustment are complex, and do not affect the results presented here given the

small amount of missing data. So, the full entire technique is not explained here, but can be found in Allison (2002).

5.1.1.2 *Creation of disposition effect scale scores*

The disposition effect scale score for each participant is the sum of their responses to all 10 items. This score has a possible range of between 10 and 50, the actual range being 18 to 43, and the sample has a mean of 31.2. All imputations are normally distributed (Kolomogorov-Smirnov test, all imputation p-values $>.100$). Figure 5.1 below shows the distribution of imputation 1, but all are similar.

Figure 5.1 Distribution of disposition effect scale scores (imputation 1)



5.2 CONVERGENT VALIDITY OF THE DISPOSITION EFFECT: CORRELATIONS BETWEEN TRADING DATA AND THE GAME

Now that disposition effect scores have been created from the scale, it is possible to compare them with DE scores from the two-index game to test convergent validity. As noted earlier, if the disposition effect is a stable bias of an individual it should be possible to measure it in multiple ways and still find consistent results.

The key strength in establishing convergent validity using these data is that the scale and two-index game are very different methods of trying to measure the same behaviour. The scale comprises self-reported responses (from memory) about typical attitudes and decisions during an investor's real-life trading. By contrast the disposition effect in the two-index game is estimated directly from trading decisions made within its simplified trading environment.

Correlations between the two measures are shown in table 5.2. (A reduced sample size, $n=96$, is because some participants did not complete all plays of the game.) These correlations are the result of pooling the correlations from each imputation, as discussed in the previous section. The scale score is correlated with plays 2-4 of the game giving 3 correlations in total. There was little data from play 1 of the Milan study (as discussed in chapter 4) so this play is omitted.

Correlations are positive for all three plays: play 3 is significant, while plays 2 and 4 are highly significant. The average effect size is just above 0.3, representing a medium effect size. These results provide substantial evidence for convergent validity of the disposition effect across different domains. They also constitute prima facie evidence that further development of the scale is worthwhile.

Table 5.2 Correlations between disposition effect scale scores and DE scores

	Play 2	Play 3	Play 4
Correlation	.332	.231	.369
p-value	<.001***	.023*	<.001***

5.3 SPLITTING THE SCALE INTO FACTORS

5.3.1 Inter-item correlations and possible groups

The section above analysed scores for the whole scale. To test whether the scale relates to cutting gains and holding losses separately it needs to be split into underlying factors. A first step in analysing the scale is to look at inter-item correlations. This can give some idea of the patterns which may be present in the data.

10 items produce 45 inter-item pairs, and 15 are significant at the 5% level.¹² Table 5.3 below shows the inter-item correlation matrix. Significant correlations are highlighted, and items are grouped together based on patterns of significant correlations. This produces one group for holding losses and one for cutting gains, with item 2 cross-loading on both.

Table 5.3 Disposition effect scale inter-item correlations

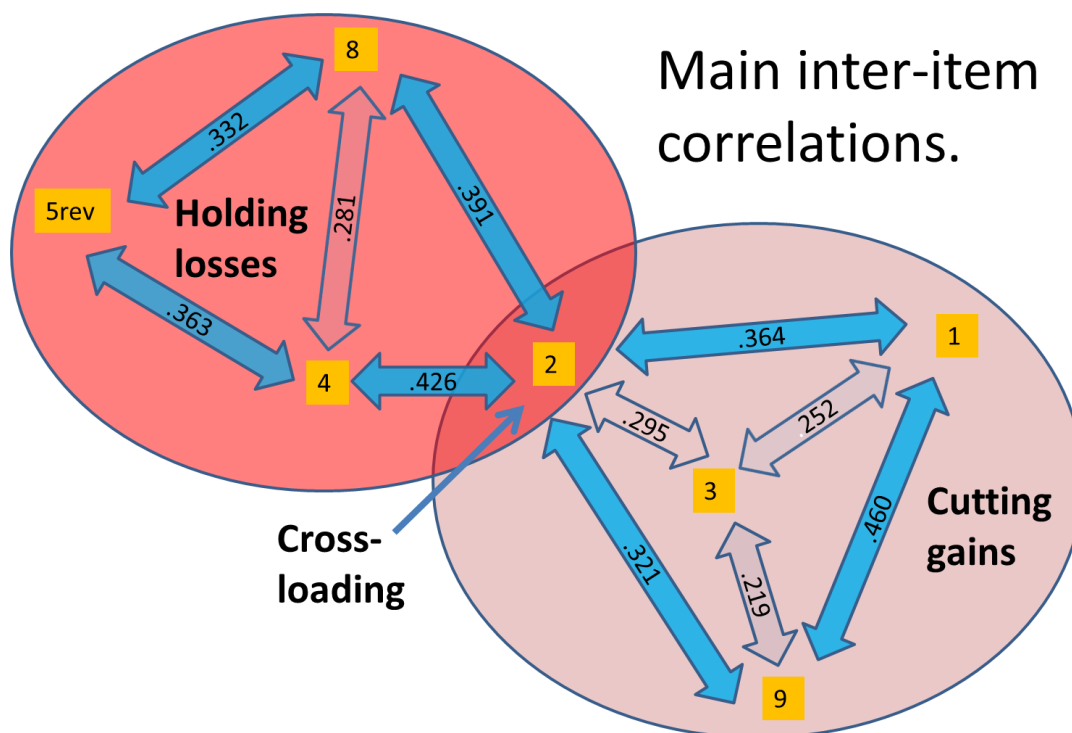
		Holding losses			Cross - loading	Cutting gains			No grouping		
		item 4	item 8	item 5	Item 2	item 1	item 9	item 3	item 7	item 6	item 10
item 4	r										
	p										
item 8	r	0.281									
	p	0.003									
item 5	r	0.363	0.332								
	p	0.000	0.000								
item 2	r	0.426	0.391	0.095							
	p	0.000	0.000	0.331							
item 1	r	0.071	0.177	0.021	0.364						
	p	0.467	0.069	0.829	0.000						
item 9	r	0.154	-0.007	-0.084	0.321	0.460					
	p	0.112	0.942	0.391	0.001	0.000					
item 3	r	0.110	0.062	-0.109	0.295	0.252	0.219				
	p	0.273	0.540	0.285	0.002	0.013	0.027				
item 7	r	-0.026	0.115	0.075	0.008	0.231	0.179	-0.025			
	p	0.791	0.237	0.444	0.932	0.016	0.065	0.804			
item 6	r	0.075	0.131	0.291	0.110	-0.056	-0.054	0.120	0.010		
	p	0.442	0.180	0.002	0.259	0.567	0.580	0.250	0.922		
item 10	r	-0.018	-0.020	0.026	-0.171	-0.071	0.046	-0.291	0.034	-0.190	
	p	0.852	0.835	0.793	0.076	0.468	0.635	0.003	0.728	0.049	

¹² Sample size was 108. After multiple imputation there were no missing data. The Bonferroni correction is not used. At a 5% confidence level and 45 correlations, about 2 false positive correlations would be expected. However, the correlations are not independent, so the appropriateness of a Bonferroni correction is arguable. The pattern of inter-item correlations provides insight into the underlying structure of the items, even if the specific values of each correlation are treated with caution due to the possibility of a few false positives.

This pattern of inter-item correlations is also shown as a Venn diagram in figure 5.2. This shows the two main groups, which include all correlations greater than $r = 0.3$ ¹³.

Correlations greater than $r = 0.3$ are shown as blue arrows, and ones below $r=0.3$ are shown as clear arrows. The overall picture is that there are two groups, one relating to holding losses and one relating to cutting gains; however, not all 10 items are part of these groupings, and item 2 unexpectedly cross-loads onto both.

Figure 5.2 Venn diagram of main groups of inter-item correlations



The largest correlation is between item 1 and 9 which both relate to cutting gains, though they are correlated with item 3 which asks about loss aversion rather than cutting gains specifically. Item 7 relates to holding gains and is reversed scored. Theoretically it should group with items 1 and 9, but its correlations are lower than

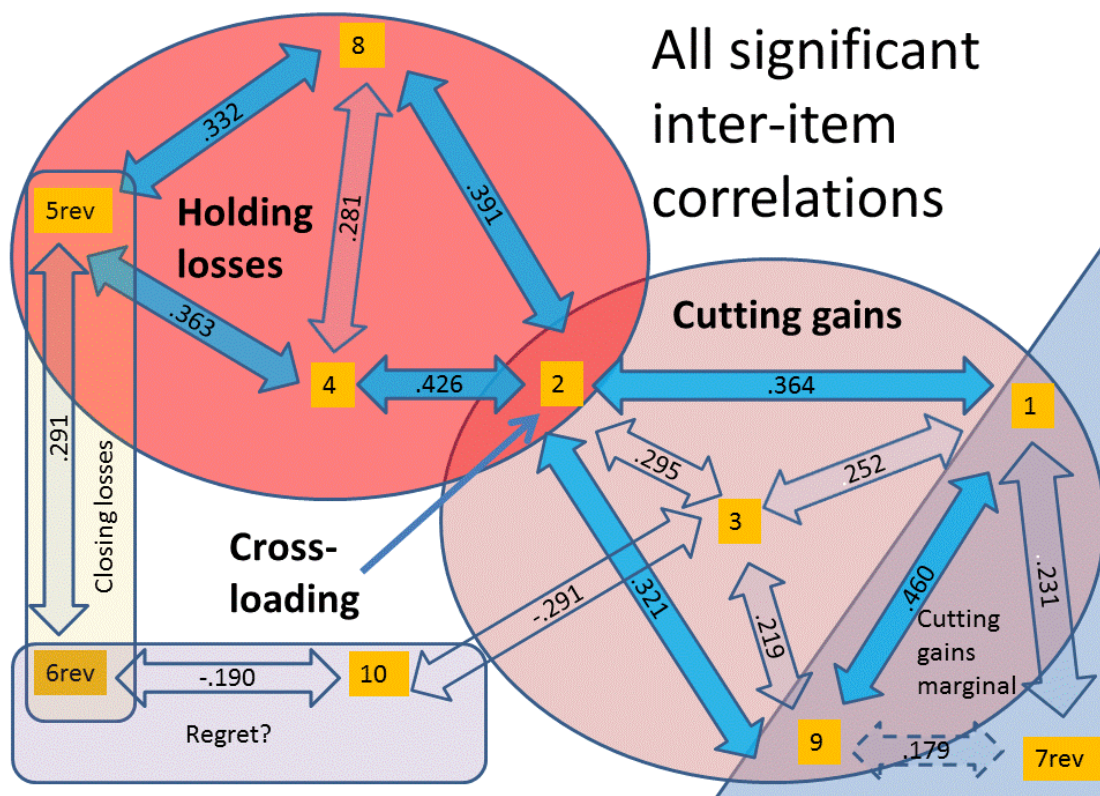
¹³ It is useful to note that the Bonferroni correction level would be 0.011 (0.05/45). With $n=108$ this p value is achieved with $r = .31$, so at this p-value the pattern of significant correlations would simply be all the blue arrows in the Venn diagram. This also assumes using the more conservative two-tailed test of correlation direction.

would be expected: it has a significant but fairly low strength correlation with item 1, and only correlates marginally with item 9.

Items 4 and 8 relate to holding losses, and correlate with each other as well as 5, a reversed item about cutting losses. Item 2 also relates to holding losses and correlates strongly with 4 and 8, though surprisingly, it also correlates with the items relating to cutting gains.

Figure 5.3 is a Venn diagram including all significant inter-item correlations ($p < .05$), not just only the main groupings. There are a small number of miscellaneous correlations, which do not form groupings into 3 or more items.

Figure 5.3 Venn diagram of all significant inter-item correlations



Items 5 and 6 correlate with each other; these items are both reversed and deal with closing losses, but item 6's only other significant correlation is a negative one with item 10. This latter correlation appears to describe how accepting losses versus forgoing gains

have opposite effects on feelings of regret: the less regret felt when closing losses (item 6), the more regret is felt when forgoing gains after closing a gain (negative item 10). In turn, item 10 correlates negatively with item 3, indicating that higher loss aversion is also associated with more regret when forgoing gains after closing a gain.

The marginal grouping of items 1, 7 and 9 discussed earlier has also been included on this diagram. Items 1 and 7 correlate, which is expected since the former asks about cutting gains and the latter is a reversed item about holding gains. However, the correlation between items 7 and 9 is not significant. It has also been shown on the diagram since it is the highest correlation which does not reach significance ($p = .065$), and the three items together make theoretical sense.

5.3.1.1 Summary of inter-item correlations

Inter-item correlations show that there are two main groupings of items: items relating to holding losses, and items relating to cutting gains. Item 2 cross-loads between the groups; this is unexpected since it relates to holding losses. There are a relatively small number of solitary correlations which do not fit into larger groups. The grouping of items 1, 7 and 9 makes theoretical sense but is only supported marginally by the correlations found.

5.3.2 Factor analysis method

Having explored possible inter-item correlations above, this section uses a series of factor analyses to extract underlying factors from the scale, dropping items from the analysis progressively to create a better factor solution. If factors for cutting gains and holding losses can be found, they may be able to explain the patterns found in responses to items in the scale. Clear factors explain the majority of variance in a subset of items, while being largely unrelated to variance in other items. This will allow the scale to be split into one or more subscales that relate to one factor each. Ideally a subscale can be created for each factor, with at least 3 items for each factor, and at least 50% of the variance in those items accounted for by the factor.

Principal components analysis is used to extract factors since this is an exploratory analysis of the scale. Varimax is used after initial factor extraction: varimax attempts to create factors which account for as much variance as possible, and again this is appropriate for an exploratory analysis attempting to extract independent factors. The analysis was carried out using the FACTOR program (Lorenzo-Seva & Ferrando, 2006). This program allows the parallel analysis method of factor retention to be used (discussed in the factor retention section below).

Since multiple imputation was used to account for missing data, there are 10 complete datasets of the scale data. To allow FACTOR to use the information from all imputations, the matrix of the pooled inter-item correlations has been calculated using SPSS, and then this data fed into FACTOR.

5.3.3 Factor loadings and item retention

Each analysis produces a factor solution with various statistics about how the factors in that solution account for the variance in item responses. Factor loadings are the equivalent of correlation coefficients between each item and each factor. An ideal situation is for each item to load strongly onto only one factor (i.e. have a high factor loading), so that each item contributes to the measurement of one factor but not to other factors (i.e. have a low factor loading). As factor loadings can be interpreted as correlation coefficients, the square of each factor loading represents the % of variance in an item which can be explained by that factor, and is analogous to R squared.

Stevens (1996) suggests this approach is only valid for factor loadings over .4 and conventionally only rotated factor loadings greater than .4 are considered when interpreting a factor. Since these loadings are analogous to correlation coefficients, this means that at least 16% of the variance in an item is shared with the factor. Loadings below this are usually ignored.

Stevens (1996) recommends that for $n=100$ (a relatively small sample), item loadings should be greater than .512 to be accepted, based on the critical values of random factor

loadings and a two-tailed confidence level of 1%. However, in the results here, only loadings lower than 0.3 are not shown. This is to identify marginal items which load strongly enough to suggest the item correlates with a factor above chance (i.e. $0.3 < r < 0.4$), but not strongly to be included in a subscale for that factor.

The following terminology is used to refer to factor loadings. Brackets indicate that an item negatively loads on a factor.

- Loads strongly - $>.600$
- Loads moderately - $>.400$
- Loads marginally - $>.300$

The aim of this factor analysis process is to identify items which cluster together and appear in combination to measure one underlying factor. So, to achieve this, items with no strong factor loadings are dropped from the analysis while those with strong factor loadings are retained. In addition, an item which loads strongly onto one factor but marginally onto another is less distinctive than an item which loads strongly on to only one factor. In trying to find subscales which distinguish between different factors, it is better to drop cross-loading items.

The communality of an item is also a criterion for whether to retain an item. An item's communality represents the amount of variance that an item shares with all extracted factors in total. It is in fact the sum of the squares of its loadings on each factor, so it is like R^2 from the point of view of the item: how much variance in the item is explained by the factors extracted.

The higher this communality the more an item's variance can be explained by the extracted factors. The lower this communality, the more variance an item has that is unexplained by any factors in the solution, which is undesirable if the aim of the analysis is to select items which can measure latent variables represented by the factors. Items with low communalities will be dropped from solutions to increase the fit of the items to the factors extracted.

5.3.4 Eigenvalues and factor retention

From each factor analysis, several factors are extracted with associated eigenvalues. As well as deciding which items to retain, there are also decisions about which factors to retain, and the eigenvalues are used to do this. The eigenvalue for each factor represents the total variance from all items combined which each factor can explain. An eigenvalue of 1 means a factor explains as much variance as one item. A factor that represents an important underlying factor explaining variance in several items will have an eigenvalue above 1.

There are various methods of selecting the number of factors to retain. The traditional method is based on a scree plot of eigenvalue size (Cattell, 1966). However, this method is only approximately accurate when the sample size exceeds 200 (Stevens, 2002), which is not the case here.

Another popular method is Kaiser's criterion, where all eigenvalues above 1 are retained (Kaiser, 1960). However, there are significant issues with this criterion (Nunnally & Bernstein, 1994). As explained above, an eigenvalue of 1 means that a factor only explains as much variance as one item; a factor with eigenvalue of 1 is no better at explaining variance than an item itself. So, these are not good candidates for factors which could represent significant latent tendencies driving responses to the items.

Instead of Kaiser's criterion, factors are selected using parallel analysis (Horn, 1965). Parallel analysis calculates the mean eigenvalues which would be generated by random data using as the same number of items, then calculates the 95% confidence intervals of each mean. When an eigenvalue is above the 95% confidence limit, it is retained; when it is below this threshold, extraction of factors ceases.

This technique of factor selection has a similar interpretation as other significance testing. If an eigenvalue is higher than its 95% confidence interval, then the chance of finding an eigenvalue at least as large as this by chance is less than 5%. The effect is

that factors are extracted more conservatively than Kaiser's criterion, so the number of false positives is reduced. The factors extracted are much more likely to be meaningful.

5.3.5 Item reduction using factor analysis

This section contains succession of factor analyses on the scale. The aim is to identify the main factors underlying responses to the items, and which items load onto which factors. Items which do not fit the factor solution well are removed until all items remaining load onto one of the remaining factors.

The reporting of each factor analysis begins with a summary of the factors extracted and variance explained. The eigenvalues of the unrotated solution are reported and parallel analysis is applied to decide how many factors to retain. The item communalities with the retained factors are shown, and then factor loadings for each item onto the rotated factors. The reporting of each factor analysis ends with a discussion of the interpretation of the factors extracted, and which items will be dropped in the next factor analysis, to try and improve the fit of the factor solution.

5.3.6 10 items

Entering all 10 items into the analysis extracts 3 factors, which together explain 54% of the total variance. The selection of the factors is shown in table 5.4. The 1st and 2nd factors are clearly greater than their 95% intervals. The 3rd is just above its interval of 1.30 and is unlikely to be a random finding ($p = .002$). The 4th factor is below the 95% confidence interval so is not extracted.

Table 5.4 Eigenvalues and factor retention for 10 items

	F1	F2	F3	F4
Factor eigenvalue	2.39	1.65	1.35	1.07
Eigenvalue 95% interval	1.65	1.45	1.30	1.19
Eigenvalue p-value	-	-	.002	>.05
% variance explained	24	16	13	

Table 5.5 Communalities for 10 items

Item	Communality
1	0.623
2	0.623
3	0.610
4	0.481
5	0.656
6	0.406
7	0.276
8	0.509
9	0.618
10	0.582

Table 5.6 Factor loadings for 10 items

Strength	All	F1	F2	F3
Good >.6		4 .666 5 .778 8 .694	1 .779 9 .785	3 .702 10 (.759)
Moderate >.4		2 .472 6 .427	2 .514 7 .406	
Marginal >.3			3 .341	6 .390 2 .369 7 (.318)
Eigenvalue	5.38	1.95	1.93	1.50
% variance	54	20	19	15

This solution reproduces the main groupings seen in the inter-item correlations. The first factor includes items 4, 5 and 8 which load strongly onto it, and so is easily identified as the loss factor. Items 2 and 6 also load moderately on to holding losses, reinforcing its interpretation.

The second factor has two items (1 and 9) loading strongly onto it. So, the bulk of this factor is about cutting gains, but there isn't a clear third item to add to make a subscale. Other items load onto it to a lesser extent. Item 7 makes theoretical sense, since that asks about holding gains (reversed item). Item 3 indicates loss aversion, while item 2 is about taking risks with losses, so this factor's interpretation is more mixed but has a strong cutting gains element.

The third factor is more difficult to interpret. It includes two strongly loading items: item 3 indicating loss aversion, and item 10 indicating **not** regretting closed items that rise further. Item 10 loads negatively, so the two items could be summarised thus: investors who feel losses more than gains (high score on item 3) tend to feel more regret when a closed gain rises further (low score on item 10).

Item 6 just misses the .4 criterion for the third factor. It asks about closing losses "without regret", and is reverse-scored. So, people who feel losses more than gains (high score on item 3) tend to feel more regret when closing losses (high score on item 6 after reverse-scoring).

Feeling more regret when closing losses (items 6) and feeling losses more than gains (item 3), are consistent with a disposition effect. However, the loading of item 10 does not seem consistent with representing the disposition effect. Item 10 loads negatively: someone with high loss aversion (high item 3) was more likely to cut gains and feel regret (low item 10). The usual explanation of the disposition effect is that investors are too eager to sell gains. If an investor felt that they were justified in cutting gains to avoid an undesirable decrease in value, they would presumably not feel regret.

Taking these three items together, the factor interpretation for the third factor is that investors who feel losses more than gains also feel more regret, both when they close losses and when they miss out on further gains. It is possible that the interpretation of item 10 in terms of the disposition effect is incorrect. For example, perhaps both incurred losses (from selling at a loss), and forgone gains (from selling a gain too early)

are both treated as if they are losses by the investor, and both generate regret associated with a loss.

Items 2 and 7 also load marginally onto this factor. However, it is difficult to draw a common thread between them and the other items. Factor interpretation should be improved by avoiding them loading onto this factor. Item 7 has the lowest communality of the 10 items by some margin at .270, i.e. only 27% of its variance is explained by the 3 factors combined. So, item 7 will be dropped, and the remaining items re-analysed in the hope that this improves the factor interpretation.

5.3.7 9 items (item 7 dropped)

Dropping item 7 still results in a 3-factor solution, explaining an improved 59% of the variance. The 1st and 2nd factors are again clearly greater than their 95% intervals. The 3rd is again just above its interval of 1.26 so is unlikely to be a random finding ($p = .005$) and is extracted. The 4th factor is below the 95% confidence interval so again is not extracted.

Table 5.7 Eigenvalues and factor retention for 9 items

	F1	F2	F3	F4
Factor eigenvalue	2.36	1.64	1.29	.85
Eigenvalue 95% interval	1.59	1.41	1.26	1.15
Eigenvalue p-value			.005	>.05
% variance explained	26	18	14	

Table 5.8 Communalities for 9 items

Item	Communality
1	0.579
2	0.644
3	0.596
4	0.525
5	0.648
6	0.514
8	0.503
9	0.620
10	0.662

Table 5.9 Factor loadings for 9 items

Strength	All	F1	F2	F3
Good >.6		4 .685	1 .775	3 .702
		5 .773	9 .785	6 .554
		8 .693	2. 596	10 (.811)
Moderate >.4		2 .473	3 .445	
		6 .368		
Marginal >.3				
Eigenvalue	5.28	1.92	1.93	1.43
% variance	59	21	22	16

Dropping item 7 leads to a clearer 3 factor solution, explaining 59% of the variance. Indeed, the total variance explained, measured by the sum of the eigenvalues (5.28), is only a little lower than the total explained with 10 items (5.38). So, a whole item has been removed, but the variance explained has only dropped by 0.1, or 10% of the variance in one item.

The interpretation of the factor is largely the same as before, although items 2 and 3 now load more strongly onto the 2nd factor, while item 6 loads more strongly onto the third factor. This improvement with item 6 reinforces its previous interpretation as an association between loss aversion and experiencing regret.

Overall, the fit is much improved from using 10 items. A good percentage of the variance is explained with the 3 factors, and the communalities for all 9 items are above 0.500.

There are 3 items strongly loading onto each factor, and no marginally loading items on any factor.

However, the aim of this chapter is mainly to find factors that match the cutting gains and holding losses biases discussed in the previous chapter. Though the emergence of the third factor is interesting, it will not be pursued here. Since item 10 is the only item which does not load onto one of the two main factors, it will be dropped.

5.3.8 8 items (items 7 and 10 dropped)

Dropping items 7 and 10 results in a 2-factor solution, explaining 49% of the variance.

The 1st and 2nd factors are clearly greater than their 95% intervals, while the 3rd factor now falls below its interval of 1.22, so is not extracted.

Table 5.10 Eigenvalues and factor retention for 8 items

	F1	F2	F3
Factor eigenvalue	2.32	1.63	1.00
Eigenvalue 95% interval	1.56	1.35	1.22
Eigenvalue p-value			> .05
% variance explained	29	20	-

Table 5.11 Communalities for 8 items

Item	Communality
1	0.563
2	0.640
3	0.326
4	0.485
5	0.628
6	0.256
8	0.486
9	0.574

Table 5.12 Factor loadings for 8 items

Strength	All	F1	F2
Good >.6		4 .645	1 .750
		5 .772	2. 661
		8 .672	3 .571
			9 .751
Moderate>.4		2 .451	
		6 .493	
Marginal >.3			
Eigenvalue	3.96	1.92	2.04
% variance	49	24	25

Although dropping item 10 simplifies the factor structure, this solution is now a poor fit, explaining less than half the variance. The communality for item 6 is particularly low (.256) so this item will be dropped and the remaining items re-analysed.

5.3.9 7 items (items 6, 7 and 10 dropped)

7 items also produce a 2-factor solution, explaining an improved 55% of the variance. The 1st and 2nd factors are clearly greater than their 95% intervals, while the 3rd factor still falls below its interval of 1.16 so is not extracted.

Table 5.13 Eigenvalues and factor retention for 7 items

	F1	F2	F3
Factor eigenvalue	2.29	1.54	.86
Eigenvalue 95% interval	1.51	1.31	1.16
Eigenvalue p-value			> .05
% variance explained	33	22	-

Table 5.14 Communalities for 7 items

Item	Communality
1	0.555
2	0.642
3	0.375
4	0.549
5	0.612
8	0.532
9	0.572

Table 5.15 Factor loadings for 7 items

Strength	All	F1	F2
Good >.6		4 .720	1 .738
		5 .743	3 .612
		8 .722	9 .756
			2. 611
Moderate>.4		2 .519	
Marginal >.3			
Eigenvalue	3.83	1.87	1.96
% variance	55	27	28

The two-factor solution is very similar to the previous one, though item 2 has a slightly higher loading than before on factor 1 (holding losses). Dropping item 6 is an improvement in terms of fit, with the variance explained increasing to 55%. This is reflected in the small change in eigenvalues total: the total has fallen from 3.96 to 3.83, so a drop of variance representing only 13% of one item, though a whole item (i.e. 100% variance of one item) has been removed from the analysis.

The main flaw with this solution is that item 2 still cross-loads on both factors. Since an aim of the analysis is to produce distinct factors, it will be dropped and the remaining items analysed, leaving 3 items per factor.

5.3.10 6 items: 4, 5, 8, 1, 3, 9

As expected, these 6 items produce a 2-factor solution, which explains a slightly improved 56% of the variance. The 3rd factor is not extracted.

Table 5.16 Eigenvalues and factor retention for 6 items

	F1	F2	F3
Factor eigenvalue	1.80	1.54	.84
Eigenvalue 95% interval	1.47	1.26	1.11
Eigenvalue p-value			> .05
% variance explained	30	26	-

Table 5.17 Communalities for 6 items

Item	Communality
1	0.622
3	0.377
4	0.537
5	0.661
8	0.519
9	0.623

Table 5.18 Factor loadings for 6 items

Strength	All	F1	F2
Good >.6		4 .720 5 .743 8 .722	1 .738 3 .612 9 .756
Moderate >.4			
Marginal >.3			
Eigenvalue	3.34	1.68	1.66
% variance	56	28	28

As expected, dropping item 2 produces two distinct factors, with 3 items each loading strongly onto them, and every item loading strongly onto only one factor. The loss factor comprises items 4, 5 and 8, and the holding gains factor comprises items 1, 3 and 9.

Holding losses is conceptually clearer than cutting gains, which includes one item about general loss aversion (item 3). Interestingly, loss aversion does not load significantly onto the loss factor. If any item was cross-loading onto both factors, it would arguably be expected to be this one, since loss aversion should theoretically be linked to both biases.

Communalities are all high, except for item 3. Although item 3 has a factor loading of .612, which is far above the .4 criterion, its communality is poor in comparison to the other items. So, this seems to indicate that it is not an ideal item to measure cutting gains with.

5.3.11 Final factors produced

Overall this iterative factor analysis has demonstrated that two factors can be extracted which are representative of cutting gains and holding losses. A single disposition effect construct does not emerge from the scale. Instead there are two biases being measured, which have discriminant validity with each other. The subscales are as follows:

Cutting gains:

- Item 1 - If a trade is succeeding I would rather close the trade than take risks for the chance of further gains.
- Item 3 - I feel the effects of losses more than gains
- Item 9 - When the value of a trade I have made rises, I always close its position before its value can fall again

Holding losses:

- Item 4 - When the value of a trade I have made falls, I am usually confident it will rise in value again
- Item 5 - When the value of a trade I have made falls, I immediately close its position (reversed-item)
- Item 8 - In trading, you need to take some risk when your trade is falling and not abandon it immediately.

The next section reinforces this evidence for discriminant validity, by further testing whether the extracted factors can be considered independent of one another.

5.4 TESTING THE INDEPENDENCE OF THE FACTORS

An aim of this chapter was to show that two factors could be extracted, corresponding to cutting gains and holding losses. There is some evidence of discriminant validity between them simply from the fact that it was possible to extract them – the two factors clearly measure different behaviours. However, this section will provide stronger evidence for

discriminant validity by testing whether the two factors are statistically independent of one another.

The iterative factor analysis above to extract them was carried out using varimax rotation. This rotation method extracts factors which are orthogonal to one another (i.e. have a correlation of zero). This is the most common rotation used in exploratory scale analysis, and tends to produce factors which are more interpretable because there will be no shared variance between them. However, a weakness of this method for demonstrating discriminant validity is that the factors extracted may have simply been forced to be independent by the varimax rotation method.

In this section, the factor analysis on the final 6 items is repeated but using direct oblimin rotation as the rotation method, which allows the factors to correlate freely. To show discriminant validity, the correlation between the factors should be as close to zero as possible.

In addition, a factor solution including a second-order factor will also be tested. If the first-order factors in the original analysis both load onto a second-order factor, this shows that some of the variance in their scores is driven by this higher-order factor. Shared variance between cutting gains and holding losses in this scale could indicate they are both driven in part by an underlying second-order disposition effect factor, contrary to the prediction.

5.4.1 6 items: 4, 5, 8, 1, 3, and 9, using direct oblimin rotation

As expected, these 6 items produce a 2-factor solution, which explains a slightly improved 56% of the variance. The 3rd factor is not extracted as its eigenvalue p-value indicates the factor occurs due to chance.

Table 5.19 Eigenvalues and factor retention for 6 items with direct oblimin rotation

	F1	F2	F3
Factor eigenvalue	1.80	1.54	.84
Eigenvalue 95% interval	1.45	1.26	1.12
Eigenvalue p-value			> .05
% variance explained	30	26	-

Table 5.20 Communalities for 6 items with direct oblimin rotation

Item	Communality
1	0.622
3	0.377
4	0.537
5	0.661
8	0.519
9	0.623

Table 5.21 Factor loadings for 6 items with direct oblimin rotation

Strength	All	F1	F2
Good >.6		4 .709 5 .799 8 .712	1 .774 3 .616 9 .791
Moderate >.4			
Marginal >.3			
Eigenvalue	3.34	1.66	1.68
% variance	56	28	28

The interpretation of this solution is very similar to the 6-item analysis carried out with varimax rotation. It still produces two factors, holding losses and cutting gains, with the same 3 items each that load strongly onto them. The crucial difference is that the correlation between these factors has been allowed to correlate freely rather than being forced to equal zero. However, this correlation is still only $r = .073$, which for practical purposes is still zero. On this basis, there is good evidence for treating these two factors as independent (Pedhazur and Schmelkin, 2013).

The results of the 2nd order factor solution are shown in tables 4.21 and 4.22. The factor loading of factor 1 onto the 2nd order factor is very weak, and its communality is almost

zero. This finding is reinforced by examining the factor loadings of the individual items onto the first and second order factors. The items in factor 1 do not load onto the 2nd order factor at all (the loadings for the holding losses items were all below 0.2, which is far below the 0.4 criterion considered to be meaningful). Adding the 2nd order factor in this solution simply splits variance of factor 2 between itself and the 2nd order factor; the factor solution would fit better by omitting the 2nd order factor, since the two first-order factors are sufficient to explain variance of the items already.

What these results demonstrate is that the two factors do not share any significant common variance: they have discriminant validity and are in effect independent of one another.

Table 5.22 2nd order factor loadings and communalities

	Loading onto 2nd order factor	Communality with 2nd order factor
Factor 1	.100	.01
Factor 2	.733	.537

Table 5.23 Factor loadings for 6 items with 2nd order factor

Strength	All	F1	F2	2nd order factor
Good >.6		4 .706		
		5 .795		
		8 .709		
Moderate>.4			1 .527	1 .578
			3 .419	3 .448
			9 .539	9 .576
Marginal >.3				

5.5 CORRELATIONS BETWEEN THE SCALE FACTORS AND THE TWO-INDEX GAME

The factor analysis of the trading scale has produced two factors, which approximately represent cutting gains and holding losses. Having extracted the two factors, it is now possible to use data from the two-index game and the scale simultaneously to test convergent and discriminant validity of cutting gains and holding losses.

To recap the logic of this testing, convergent validity is demonstrated where variables which are theoretically related to each other, are found to be related to each other when measured. Convergent validity provides evidence that a bias has trait-like characteristics, because it shows the same proposed construct can be measured in different ways producing consistent results.

Discriminant validity is the opposite quality to convergent validity: variables which are not theoretically related to each other should not be related to each other when they are measured. Applied here, discriminant validity would show that cutting gains and holding losses are distinct from one another when measured, and ideally are independent of one another.

Convergent validity of the disposition effect has already been tested earlier in this chapter, by comparing DE scores from the game with overall scores on the scale. The same technique is now applied to cutting gains and holding losses, by comparing each bias measured by the game (by PGR and PLR respectively) with each bias measured by the scale (gain factor and loss factor scores respectively).

If convergent validity holds then PGR will correlate positively with the gain factor, and PLR will correlate negatively with the loss factor. For cutting gains, higher scores in both PGR and the gain factor indicate a greater tendency to sell gains, so a positive correlation is expected. The reason for the negative correlation between the loss factor and PLR is that a high loss factor score represents a strong tendency to hold onto losses longer. An individual who holds losses longer will have a lower PLR score, not a higher one. So, the bottom half of the PLR fraction will be larger since it measures the duration of time spent holding losses, and this results in lower PLR scores, not higher ones.

As discussed earlier with the disposition effect, comparing scores from the game and the scale is a powerful method to demonstrate convergent validity because the two methods are very different ways to measure the same bias. The scale asks for self-reported personal reflections on usual trading patterns in real financial markets, while the two-

index game directly measures trading patterns from a brief period of trading in a simulated stock market.

Discriminant validity has already been tested within the game and within the scale. The second half of chapter 4 showed that PGR and PLR are largely independent. Likewise, the factor analysis above shows that the gains and losses factors are independent of each other. Further evidence for discriminant validity can be provided by comparing scores for the two biases across the two methods of measurement. The predictions for discriminant validity are that PGR does **not** correlate with the loss factor, and PLR does **not** correlate with the gain factor. Ideally the correlations should be as close to zero as possible.

5.5.1 Correlations between factor scores and game scores

5.5.1.1 Calculation of factor scores

To correlate PGR and PLR with scores from the scale, participants need to be assigned a score for each factor based on their responses to the scale. Two different methods are used to create these scores. The first is factor scores: these represent each individual's score on the latent factor which the factor analysis has extracted. These scores are calculated by multiplying the factor loadings for each item by the responses participants gave for each item. So, it captures how important each item is to a factor, and whether a respondent scored low or high on each item. The effect is conceptually similar to using a weighted average model. The regression method has been used, as this method allows the scores for different factors to correlate.

These scores were produced using SPSS. Since SPSS does not have the functionality to produce these factor scores across multiple imputations, factor scores were created for all 10 imputations. The correlations reported below are the results for the pooled results of all 10 imputations. Correlations are calculated using a sample size of 96.

5.5.1.2 Testing convergent validity

Results for the convergent validity of cutting gains show no significant correlations between PGR scores and the gain factor, which is unexpected. The correlation with PGR2

is marginal; the three correlations could be taken as a group to show a trend of least being positive rather than clustering around zero, however, the effect sizes are too small to approach statistical significance, so this is not strong evidence for this conclusion. These results are explored further later in this chapter.

Table 5.24 Correlations between PGR scores and gain factor scores

	PGR2	PGR3	PGR4
Correlation	.189	.121	.050
p-value	.065	.243	.629

In contrast, results for the correlations between PLR scores and loss factor scores are much stronger, and confirm what was predicted. The loss factor correlates significantly with every play of PLR, and the average correlation coefficient over the three plays is just under $r = -0.3$, which indicates a medium-sized effect. So, this supports convergent validity between the game and scale for holding losses.

Table 5.25 Correlations between PLR scores and loss factor scores

	PLR2	PLR3	PLR4
Correlation	-.300	-.242	-.332
p-value	.003**	.017*	>.001***

5.5.1.3 Testing discriminant validity

Testing discriminant validity between the gain factor, and PLR scores are more consistent with the hypothesized relationship. There are no significant correlations, as expected. In fact, not only are these correlations all significant, but they appear to cluster around $r = 0$, suggesting the associations between them is completely random. Knowing about a participant's PGR score tells us nothing about their likely loss-factor score.

Table 5.26 Correlations between PGR scores and loss factor scores

	PGR2	PGR3	PGR4
Correlation	.003	.064	-.083
p-value	.974	.540	.425

The failure to find convergent validity between PGR scores and gain factor scores suggests that this factor is not been well represented by the scale. However, there is still some value in testing for discriminant validity between the gain factor and PLR scores. Although PLR correlates with the loss factor, and the gain and loss factors are independent, it is still possible that PLR is related to both the gain factor as well. The gain factor could measure some aspects of holding losses.

The results show that this is not the case. As with the correlation between the gain factor and PLR scores, the correlation coefficients cluster around zero suggesting random association between the two. Knowing a participant's PLR score tells us nothing about their likely gain-factor score.

Table 5.27 Correlations between PLR scores and gain factor scores

	PLR2	PLR3	PLR4
Correlation	-.061	.041	-.024
p-value	.559	.694	.818

Combining the two tests in tables 5.26 and 5.27, there is strong evidence for discriminant validity between cutting gains and holding losses, when measured with the scale and the game.

5.5.2 Correlations between scale scores and game scores

An alternative method is to identify the items which load strongly onto each factor, and sum the responses to those items. So, the sum of items 1, 3 and 9 is the gain factor score, and the sum of items 4, 5 and 8 is the loss factor score. For clarity, these totals are referred to as scale scores, rather than factor scores. They may measure the scores on the latent factors less accurately, because they treat the 3 items in each cluster as equally important (in effect each item is given a 33% weighting towards its factor). However, they can be seen as a robustness check on the process of creating factor scores. Again, correlations are calculated using listwise deletion and sample size of 96.

5.5.2.1 Testing convergent validity

Results using the scale scores are similar to those using factor scores. Cutting gains correlations are still non-significant, while holding losses are still significant, except for PLR3 which is now only marginal. However, the overall conclusions are much the same.

Table 5.28 Correlations between PGR scores and gain factor scale scores

	PGR2	PGR3	PGR4
Correlation	.155	.091	.022
p-value	.133	.382	.830

Table 5.29 Correlations between PLR scores and loss factor scale scores

	PLR2	PLR3	PLR4
Correlation	-.266	-.190	-.360
p-value	.009**	.064	>.001***

5.5.2.2 Testing discriminant validity

Again, the results are very similar to those using factor scores. The cross-correlations between different biases are all close to zero. So, the evidence for discriminant validity, and indeed independence between biases, remains very strong.

Table 5.30 Correlations between PGR scores and loss factor scale scores

	PGR2	PGR3	PGR4
Correlation	-.031	.045	-.106
p-value	.767	.666	.310

Table 5.31 Correlations between PLR scores and gain factor scale scores

	PLR2	PLR3	PLR4
Correlation	-.060	.031	-.035
p-value	.561	.763	.734

5.5.3 Convergent validity measuring cutting gains

There was good evidence for convergent validity between the scale and TIG when measuring holding losses, and for discriminant validity when measuring cutting gains and holding losses. However, the evidence for convergent validity when measuring cutting gains was weak. This section discusses this finding, and includes further analysis to investigate it. For reference, the items from the scale relating to gains are reproduced again in the table below.

Table 5.32 Items from the scale relating to gains

#	Item text	Gain/ Loss	Trading pattern or attitude
1	If a trade is succeeding I would rather close the trade than take risks for the chance of further gains.	Gain	Cutting gains
3	I feel the effects of losses more than gains	Both	Loss aversion
7	When the value of a trade I have made rises, I wait for it to drop before closing its position.	Gain	Holding gains (reversed item)
9	When the value of a trade I have made rises, I always close its position before its value can fall again	Gain	Cutting gains
10	When the value of a trade I have closed rises still further, I accept the situation without regret.	Gain	Cutting gains

The correlations were all positive as expected, however they were not large enough to reach significance. The smaller correlations may be caused by too much noise in the measurement of cutting gains by either the two-index game or the scale, or the scale may not measure cutting gains accurately. The game has been shown to reliably measure PGR already; therefore, it is more likely that a failure to find significant correlations is due to issues with the gain factor in the scale.

5.5.3.1 Issues with item 3 in the gain factor

The gain factor was extracted from a small sample of scale items. Since there was not a wide selection of items from which to pick the best fitting items, this may have limited how well the items included in the gain factor (items 1, 3 and 9) measure cutting gains. While the loss factor has three items which correlate closely with each other, and all had very strong factor loadings, item 3 did not fit as well with items 1 and 9 in the gain factor.

Indeed, *prima facie* item 3 appears to be about loss aversion. Someone who feels “the effect of losses more than gains” should be more likely to hold losses (avoiding the negative emotions from selling losses) than cut gains (experiencing the positive emotions from selling gains). Item 1 refers to selling gains while forgoing the prospect of further gains, and item 9 refers to selling gains in order not to lose the gains made. So, the positive correlation between items 1 and 9 with item 3 could be interpreted as some participants thinking about their situation when already holding a gain, and weighing a future further gain versus a reversal of the gain they are holding. This would explain the non-significant correlations of item 3 with the holding losses items (4, 5 and 8). With a simpler interpretation of item 3 as representing loss aversion, item 3 was expected to correlate with holding losses.

While the iterative factor analysis above showed that item 3 was the best item available, this discussion above does suggest that a future scale would benefit from developing from a wider range of initial items to refine measurement of cutting gains and be more specific about what situations participants are referring to. In particular, the relationship of item 3 to the other gain items, and item 3’s lack of correlation with holding losses, need to be studied in more detail.

5.5.3.2 Gain factor possibly composed of distinct behaviours

Given the issues with item 3, another interesting avenue to explore the failure to find a clear cutting gains factor from the scale involves omitting item 3, and also hypothesising

that cutting gains is not a unitary bias either, but has distinct behaviours that contribute to it. So, it would be like the disposition effect, which is the combination of cutting gains and holding losses. These constituent behaviours may be independent or close to independent of each other; if so items measuring those behaviours would not correlate with each other, even though they both correlate with the cutting gains factor overall.

In the 10-item scale there are 4 items which specifically relate to trading behaviour towards gains: 1, 7, 9 and 10. (These all related to cutting gains except 7, which was a reversed item about holding gains.) Table 5.33 shows correlations between PGR and sum of these four items, which I will refer to as the 'all-gains scale score'. Using this theory-driven selection of items to produce scale scores produces some evidence for convergent validity between the scale and PGR. One correlation is now significant, and the other two are marginal.

Table 5.33 Correlations between PGR scores and all-gains scale scores

	PGR2	PGR3	PGR4
Correlation	.178	.220	.178
p-value	.083	.031*	.083

Continuing this theory-based approach, items that relate specifically to selling gains (items 1, 9 and 10) could be grouped together. I will refer to this as the 'selling-gains scale score'. Table 5.33 shows correlations of these scores with PGR. This new subscale has reasonable correlation with PGR: all three correlations are significant. The average effect size is around $r = .250$, lower than for holding losses, but still a medium effect size. Although this grouping is not supported by the iterative factor analysis earlier in this chapter, the theoretical rationale for grouping these items is a clear one. It seems unlikely to be a coincidence that the three items that relate specifically to selling gains do in fact produce significant positive correlations with a behavioural measure of selling gains (PGR). So, the evidence for construct validity of cutting gains is not strong, but it is not completely absent.

Table 5.34 Correlations between PGR scores and selling-gains scale scores

	PGR2	PGR3	PGR4
Correlation	.238	.278	.207
p-value	.019*	.006**	.043*

If the correlations between the selling-gains scale score and PGR are evidence of a real connection, it raises the question why the three items included do not group together in the factor analysis. As mentioned above, one possibility is that cutting gains is itself not a unitary bias. Perhaps all three items measure some constituent part of cutting gains. For example, items 1 and 9 may measure the behaviour 'cutting gains to avoid risk', while item 10 may measure the behaviour 'cutting gains without regret'. The result would be that all three items correlate with PGR, because the proposed constituent behaviours would also both correlate with PGR. However, items 1 and 9 would not correlate with item 10, because they measure separate behaviours which are largely independent of each other.

This situation would be analogous to the disposition effect itself and its two constituent biases, as argued in this chapter. Items for both cutting gains and holding losses will correlate with the disposition effect, because they both measure stable aspects of the disposition effect. However, they do not correlate with each other because they measure separate behaviours that individually contribute to the disposition effect.

The correlation between PGR and the single item 10 is obviously not sufficient for a useable scale. In the case of item 10, its failure to form a strong cluster of items relating to cutting gains and regret could be the absence of related items in the original scale. If the disposition effect comprises at least 3 constituent biases (two for cutting gains and one for holding losses), then only 10 initial items were not ideal for exploratory factor analysis. However, the results here suggest that further development of the scale, and especially the gain factor, could be fruitful. More initial items from which to extract theoretical factors looks like a promising avenue to develop the scale and better define possible constituent biases within the disposition effect.

5.6 CONCLUSION

This chapter has aimed to build on the results from chapter 4, by introducing the disposition effect scale. Combining this data with data from the two-index game, more extensive testing of convergent and discriminant validity has been carried out on the disposition effect, cutting gains and holding losses.

Convergent validity for the disposition effect was successfully demonstrated. DE scores and disposition effect scale scores correlate significantly, with an average effect size of roughly $r = 0.3$, a medium sized effect. This supports the disposition effect as a stable behavioural tendency which drives both variables, and that the disposition effect scale is effective at measuring the disposition effect in investors.

There was also significant evidence for convergent and discriminant validity when measuring holding losses, and some evidence for them when measuring cutting gains. Two factors were extracted which appear to approximately represent cutting gains and holding losses. These factors were also found to be independent of one another, providing evidence of discriminant validity between them.

Testing these factors in combination with the two-index game, the loss factor correlated significantly with PLR, again with an effect size of approximately $r = 0.3$, demonstrating convergent validity. Crucially the loss factor did not correlate with PGR, demonstrating discriminant validity. Results for the gain factor were mixed. There was discriminant validity between the gain factor and PLR; however, there were no significant correlations between the gain factor and PGR.

To investigate this unexpected result, alternative gain factor scale scores were calculated by combining item scores based on theory rather than the outcome of the factor analysis. A theory-based 'selling-gains scale score' did have significant correlations with PGR scores. Why this theory-based scale did not emerge from the factor analysis is an open question, though one possibility is that cutting gains is composed of several constituent biases itself. These results suggest further work in developing a longer scale,

to allow the factors extracted to be refined and capture these behaviours more accurately.

In summary, it has been demonstrated that the disposition effect scale scores have convergent validity with the DE scores. Two factors have been extracted from the scale, which represent the cutting gains and holding losses, and these have discriminant validity (in fact they are independent) from each other. Convergent and discriminant validity was also demonstrated between the scale and the two-index game, except for convergent validity between the gain factor and PGR. However, a theory-based gain scale does produce positive correlations with PGR; this suggests that further research on cutting gains may be able to produce a factor structure for cutting gains which does have convergent validity with PGR scores.

6 THE EFFECT OF REAPPRAISAL ON THE DISPOSITION EFFECT IN RETAIL INVESTORS

This chapter presents the experimental results from the Milan study, which used cognitive reappraisal as an intervention to change the level of trading biases displayed by participants. It relates to the third research question: "Does cognitive reappraisal affect the disposition effect and its constituent biases, when tested in experienced traders under conditions of greater external validity?"

The aim to test cognitive reappraisal with greater external validity is motivated by a desire to build on previous experimental work, which has shown that cognitive reappraisal can reduce trading biases including the disposition effect, but has only done so with unrealistic simplified trading tasks, and usually with student samples. The Milan study uses the two-index game, which more accurately reflects the trading environment that investors make decisions in on financial markets. In addition, participants comprised a sample of retail investors, who more accurately represent those who trade in financial markets where the disposition effect has been observed.

As seen in previous chapters, the two-index game produces individual scores of cutting gains (PGR) and holding losses (PLR) for each play of the game. This allows the separate effects of cognitive reappraisal on these two constituent biases to be tested too, also with conditions of greater validity. Testing them separately allows inferences to be made about why the disposition effect is affected: whether by a change in cutting gains, or holding losses, or both. So, this study aims to test three hypotheses relating to the third research question, in an experimental setup, but under conditions of greater external validity:

- Investors will show a disposition effect in the two-index game
- Cognitive reappraisal will reduce the disposition effect
- Cognitive reappraisal will reduce holding losses but not affect cutting gains

The chapter begins describing the method used in the Milan study, and the techniques used in data analysis. Some of this detail has been covered in chapter 3, but details specific to the Milan study are covered in this chapter. The overall disposition effect displayed by participants is tested first, and then the effect of cognitive reappraisal is examined on each bias. The effect of cognitive reappraisal on the disposition effect is tested in two ways. First, by testing the change in scores before and after cognitive reappraisal is applied. Then, a marginal model is run with data from three plays of the two-index game.

6.1 METHOD OF THE MILAN STUDY

6.1.1 Experimental design & reappraisal intervention

The aim of this study was to test the effect of applying a cognitive reappraisal instruction. Comparing scores before and after the instruction, and comparing reappraisers with a control group that simply repeated the play, would produce greater statistical power for the test. So, a repeated measures design was adopted. In addition, scores on the two-index game are susceptible to learning effects. People usually need to play it a little before they get used to the mechanics of the game. To allow for this, participants were asked to play the game four times, with the cognitive reappraisal instruction given between the third and fourth plays.

Play 1 was completed in person at a trade fair; the main value of play 1 was to develop familiarity with the game (and it may also have helped as a recruitment tool.) Difficulty playing the game could create additional noise in scores, threatening to mask any experimental effect. The game can be difficult at first: players perceive the game as moving very quickly with much player attention focussed on understanding how to play the game. This includes aspects such as understanding game mechanics, and working out a trading strategy using the predictor index.

One consequence of this could be that participants are unable to express their disposition effect, because their attention is directed elsewhere. Since the disposition

effect is a difference between decisions about gains versus losses, a necessary condition for the disposition effect must involve being aware of the current gain or loss, when a trading decision is made. If attention is directed elsewhere the measured disposition effect may not occur, or could be smaller than it would otherwise be. So repeated plays of the game gave participants more time to become familiar with the game, and hopefully express any trading biases they are susceptible to.

The remainder of participation in the study was online, and participants completed plays 2-4 through personalised links sent by email. Between play 2 and play 3, participants completed the disposition effect scale, which was analysed in the chapter 5. Groups were randomly assigned when participants signed up to the study, and the sole difference between groups was that the experimental group received a cognitive reappraisal instruction between plays 3 and 4, while the control group did not.

As outlined in chapter 2, the theoretical basis for this intervention is the proposal that the disposition effect is caused by emotions experienced during trading. Cognitive reappraisal is a form of emotion regulation which has been shown in other studies to reduce decision-making biases, and to reduce biases linked to negative emotions. Therefore, the prediction is that cognitive reappraisal will reduce the disposition effect by reducing emotions while making trading decisions.

The reappraisal instruction used in the Milan study asked participants to imagine they were an investment manager, trading on behalf of clients. This reappraisal instruction was based on those used Lee *et al.* (1998) and Sokol-Hessner *et al.* (2009), which used cognitive reappraisal to reduce trading biases (in the former, the disposition effect itself). For more details of the instruction given, and the control version, please refer to appendix 3.

6.1.2 Participant recruitment

Participants were recruited at an Open University stand at the Trading Online Expo event. This was held at Borsa Italiana (Italian stock exchange) in Milan. The main

purpose of the stand was to recruit participants to sign up to complete the study online, and the researchers on site engaged potential participants in conversation about the study. The researchers on site comprised Paul Grayson, Mark Fenton-O’Creevy, and Ben Hardy. Prof Fenton-O’Creevy spent several hours in talks and presentations at the Expo, where he discussed the links between trading and emotions, and promoted the study.

After giving consent to take part, participants filled out their personal details and took a video-based tutorial of how to play the game. All instructions were supplied in both English and Italian, which had been professionally translated. Following this they played the game for the first time (play 1). Playing the game was intended to engage their interest in the study, as well as allowing participants to become more familiar with how the game works, as discussed earlier in this chapter.

Ideally participants would have completed the whole study on site, in a controlled environment where they could be overseen by the researchers; however, it was felt impractical to do this for several reasons. First, the experimental protocol required some researcher time to give and collect forms, check that participants understood the tutorial, deal with any questions, deal with internet connections not working, social aspects of dealing with people arriving and leaving, etc. To run enough people through the study would have required several additional researchers, which were not available.

Second, participation in the study (including online) took about an hour in total. This would have increased demand for using laptops, which again may not have been possible to meet. Related to these, recruitment tended to peak during the middle of the day, so researcher time and laptop time over the two days was better spent engaging and recruiting as many participants as possible during this period, rather than facilitating a smaller number of participants through the whole study.

Third, the stand was next to the main thoroughfare of the exhibition. This was good for advertising the study, but not ideal for controlled experimental condition. Finally, at the

Expo there is a full programme of talks and events on each day, often running concurrently. Attendees have this in advance and presumably plan to attend certain events each hour. Some attendees will also have paid for courses which run for part of the day. In addition, there are many other stands for attendees to engage with. Asking participants for an hour of their time is off-putting for many attendees under these circumstances, and was a problem encountered in previous field research with the xDelia project.

After completion of the study, participants were emailed feedback on their performance and comparison to the group overall, which had been offered as an incentive to take part in the study. They were also debriefed on the disposition effect, and their own disposition effect as measured in the game.

6.1.3 Data analysis

Data are analysed in three ways. First, the overall disposition effect during the study is tested, to establish whether trading during the game produced a disposition effect. One-sample t-tests are used to test the disposition effect observed, against the null hypothesis that there is no disposition effect.

Second, the effect of cognitive reappraisal was tested with the change in trading biases between plays 3 and 4. This produces "change scores" for DE, PGR and PLR, which are tested with independent t-tests using group membership as the independent variable.

Finally, marginal models are used to test the effect of cognitive reappraisal using data from plays 2-4. Marginal models are a form of multilevel (hierarchical linear) modelling. These statistical models allow data to be analysed in terms of several levels of analysis. With this study, each play generates an observation for each trading bias. The play is the repeated measure and is the first level of analysis. These plays are clustered together at the second level, which in this study is the participant.

Multilevel modelling allows complicated models to be tested, but the marginal models here are relatively simple, since they comprise only two levels and have no random factors. However, mathematically a marginal model still has several advantages to using ANOVAs.

Multilevel models produce estimates of the individual parameters in the model, in addition to the overall F tests that ANOVAs supply. In this design, taking the control group's play 2 as the baseline, there are parameters for the effect of: play 3, play 4, reappraisal group membership, and the interactions of play 3 and play 4 with reappraisal group membership. The effect of cognitive reappraisal is specifically tested by the interaction between the reappraisal group and play 4.

Multilevel models use maximum-likelihood as an estimation method, rather than OLS (ordinary least squares). This allows participants with missing data to be included in the analysis, whereas ANOVAs require complete data. The sample size for the Milan study was not large initially, so data loss from dropping participants is undesirable. Marginal models allow the maximum number of participants to be retained. Maximum likelihood estimation also makes marginal models more robust to the effects of unbalanced groups, which is also a feature of this dataset.

Finally, multilevel models allow much greater flexibility in choice of covariance structures for repeated measures. The covariance structure describes the variance of scores on each repeated measure (i.e. play 2, play 3, and play 4) and the covariance between them. Traditional statistical methods have only two options available. A repeated-measures ANOVA assumes compound symmetry; compound symmetry means the variance of each play is expected to be equal, and all correlations between plays are equal. Alternatively, a MANOVA assumes the covariance structure is unstructured. This is like deciding *a priori* that the variables are unrelated to each other, and whatever covariance parameters happen to occur in the data are adopted in the statistical model. In other words, an unstructured covariance matrix used in a MANOVA fits the covariance

parameters to the data post-hoc. These two models are not always appropriate for the data being analysed.

6.1.3.1 *Missing data*

A small number of participants failed to make any sales for losses on some plays of the game. Unfortunately, this produces a PLR of zero, which cannot be log transformed. All these participants were verified as having held losses for significant amounts of time during the simulation, so they all had the opportunity to sell losses, but chose not to do so. Since this indicates a very strong tendency to hold losses, rather than an absence of data, these participants were given a PLR score two standard deviations lower than the mean on that play. All participants sold at least one gain, so missing data issues did not arise when calculating PGR.

6.2 OVERALL DISPOSITION EFFECT

Table 6.1 shows the mean disposition effects, on each play and overall. (Play 1 was not analysed because an IT meant the sample study was greatly reduced on this play. In addition, this play was mainly to familiarise participants with the game, as discussed above). Each mean was tested against a null hypothesis of zero. A significant disposition effect was observed on play 3, play 4, and overall. Since the DE scores from the two-index game have been log transformed, a score 0 corresponds to a PGR/PLR ratio of 1 once exponentiated. Although participants did not show a disposition effect on play 2 overall, this is consistent with participants gradually becoming more comfortable with the game, and paying more attention to the gain/loss information as they did.

Table 6.1 also shows the equivalent PGR/PLR ratio once the mean DE has been exponentiated. The DEs found on plays 3 (2.16) and 4 (1.97) indicate that participants were on average around twice as likely to sell gains as they were losses. These are within the range normally found in other studies, which tend to find a ratio between 1.5 and 2. Although play 2 was not significantly different from zero, its mean ratio of 1.29 is not too far from the range of disposition effects usually found in field research.

Table 6.1 Mean disposition effect overall and on separate plays

	N	Mean	Std. Dev	Std. Error	PGR/PLR ratio	t	df	p-value
plays 2-4	151	.246	.643	.052	1.76	4.71	150	>.001***
play 2	51	.111	.536	.075	1.29	1.482	50	.145
play 3	50	.335	.735	.104	2.16	3.225	49	.002**
play 4	50	.295	.634	.090	1.97	3.285	49	.002**

6.3 THE EFFECT OF REAPPRAISAL USING PLAYS 3 AND 4

The aim of the Milan study was to test whether cognitive reappraisal produced a change in the trading biases (DE, PGR, PLR) using a control group as a comparison. To do this, change scores have been calculated using play 3 (pre-intervention) and play 4 (post-intervention). These are shown in table 6.2.

Table 6.2 Changes in trading biases by group

Change in..	Group	N	Mean change	Std. Dev.	Std. Error
DE	reappraisal	16	-.271	.582	.146
DE	control	33	.076	.400	.070
PGR	reappraisal	16	-.051	.279	.070
PGR	control	33	.036	.258	.045
PLR	reappraisal	16	.221	.476	.119
PLR	control	33	-.040	.462	.080

Cognitive reappraisal was expected to reduce the disposition effect, so the change in DE is expected to be negative in the reappraisal group, compared with the control group. Furthermore, this change was expected to be due to a decrease in holding losses, so PLR should increase in the reappraisal group relative to the control group, while the change in PGR was not expected to differ between the groups. An increase in PLR indicates more losses are being sold, and the tendency to hold losses is decreasing. (Once logged, DE equals the logged PGR minus logged PLR. So, an increase in PLR narrows the difference between PGR and PLR, and reduces DE).

Independent t-tests for each bias are reported in the sections below, using group membership as the independent variable. The tests for DE and PLR are one-tailed, since these directions were predicted in advance, while the test for PGR is two-tailed.

6.3.1.1 *The effect of cognitive reappraisal on the disposition effect*

DE decreases in the reappraisal group while increasing slightly in the control group, and an independent t-test finds that this difference is significant, presented in table 6.3. This result confirms the hypothesis that cognitive reappraisal would decrease the disposition effect. This is medium effect, $r = .34$, Cohen's $d = .71$, that explains 11% of the variance in changes in scores.

Table 6.3 T-test of mean change in DE by group

Change in...	t	df	p-value (1-tailed)	Mean Difference	Std. error of difference
DE	-2.151	22.1	.021*	-.347	.161

Levene's test for equality of variances was significant for the change in DE ($F=6.125$, $p=.017$), so the t-test is reported without assuming equality of variances. The unadjusted result was $t(47) = -2.45$, $p=.009$). Effect sizes have been calculated based on the t-test unadjusted for Levene's test. Reduced degrees of freedom in the adjusted test results in larger effect sizes, so the conservative option has been taken. Levene's test was not significant for the changes in PGR and PLR, tested below.

6.3.1.2 *Interpreting changes in DE*

In the previous section testing overall disposition effects, the mean DE was exponentiated to transform it back to its original units, so that it can be interpreted. The mean change in DE in each group is shown in table 6.2; however, this also needs to be exponentiated to be interpreted. Once exponentiated, the mean change between plays represents a multiple applied to each group's prior mean (on play 3), in order calculate their mean on play 4. The reason this is a multiplier rather than an absolute amount added or subtracted is that figures which are added or subtracted when logged become multiplied or divided once exponentiated.

So, the mean change for the reappraisal group was $-.271$. Once exponentiated this produces a figure of $.54$, or 54%. This represents the multiplier applied to the previous

DE for the reappraisal group. This means that on play 4, the DE of the reappraisal group was on average 54% of DE on play 3, which is a decrease of 46%.

The mean change for the control group was .076. Once exponentiated this produces a figure of 1.19, or 119%. So, DE on play 4 for the control group was on average 119% of DE on play 3, an increase of 19%.

The mean difference between groups in table 6.3 shows the relative change between groups. The reappraisal group is lower by -.347. Exponentiated this is 0.45, or a multiplier of 45%. This represents the effect of being in the reappraisal group, compared to what would be expected to happen if the same participant had been in the control group. Relative to control group, the reappraisal group's DE on play 4, with reappraisal, was only 45% of their DE on play 3, before reappraisal.

To explain how this is produced, we can look at the relative changes for each group separately. The control group's DE on play 3 is multiplied by 1.19 (19% increase) to produce their play 4 DE. While the reappraisal group's DE on play 3 was multiplied by 0.54 (46% decrease) to get their DE on play 4. 0.54 as a percentage of 1.19 is equal to 0.45, which corresponds to the 45% multiplier above. This is a relative decrease of 55% for the reappraisal group, compared with the control group.

In slightly less technical language, the average relative effect of being in the reappraisal group is that a participant's DE would be 45% of what it was before the reappraisal intervention, a relative reduction of 55%. While this is not an elimination of the disposition effect, a relative reduction of 55% is an amount of practical significance.

This technique can also be applied to the confidence intervals for the difference between groups, to produce upper and lower limits of the size of the multiplier. 95% confidence intervals for the mean difference between groups in the t-test in table 6.3 are -.062 and -.630. Exponentiating the confidence intervals produces multipliers of .87 and .23. These represent relative reductions of between 13% and 77% respectively for the effect of being in the reappraisal group compared with the control group.

6.3.1.3 *The effect of cognitive reappraisal on cutting gains*

PGR falls in the reappraisal group, while increasing in the control group. However, the differences are smaller than for DE. Table 6.4 shows the independent t-test of this difference between groups, which is not significant. So, reappraisal does not reduce cutting gains, and therefore reappraisal does not reduce the disposition effect here by reducing the tendency to cut gains.

The change is the in right direction to reduce the disposition effect. A reduction in PGR means a participant was less likely to sell gains, so had a reduced tendency to cut gains. This is what would be expected to bring PGR and PLR closer together and reduce DE to zero. However, in this case the effect size was small, Cohen's $d = .31$ and $r = .15$, explaining only 2.3% of the variance in PGR change. This effect was not large enough to reach statistical significance as discussed.

Table 6.4 T-test of mean change in PGR by group

Change in...	t	df	p-value (2-tailed)	Mean Difference	Std. error of difference
PGR	-1.07	47	.291	-.086	.083

6.3.1.4 *The effect of cognitive reappraisal on holding losses*

PLR increases in the reappraisal group as expected, while decreasing slightly in the control group. The independent t-test reported in table 6.5 shows that this difference between groups is significant. This represents a medium size effect: Cohen's $d = .54$, $r = .26$, explaining 6.8% of the variance in changes in PLR scores.

This result confirms the hypothesis that cognitive reappraisal would reduce holding losses. In addition, we can conclude as predicted that the change in DE scores associated with reappraisal is mainly driven by differences in PLR, rather than differences in PGR. The reduction in the disposition effect seen earlier is because participants are more likely to sell losses.

Table 6.5 T-test of mean change in PLR by group

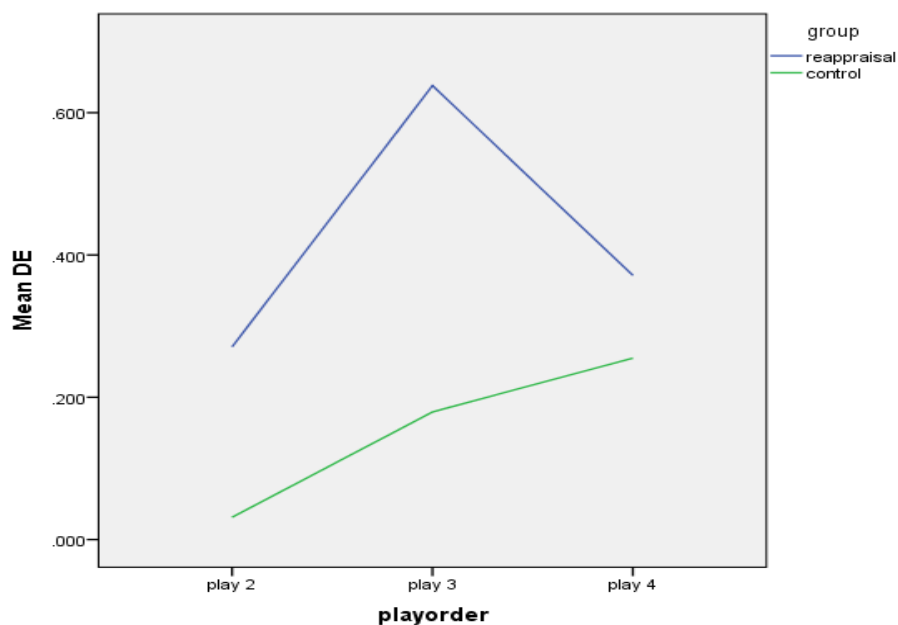
Change in...	t	df	p-value (1-tailed)	Mean Difference	Std. error of difference
PLR	1.84	47	.036*	.261	.083

6.4 THE EFFECT OF REAPPRAISAL USING PLAYS 2-4

6.4.1 DE tested over plays 2-4

The t-tests in the previous section used measurements of the disposition effect before and after reappraisal was implemented. However, analysing the data from play 2 as well, using absolute scores rather than change scores, gives a more complicated picture as seen in the figure 6.1.

Figure 6.1 Absolute DE scores over plays 2-4, split by group



It is clear that the reappraisal group has a reduction in DE on play 4, compared with the control group. However, the reappraisal group's DE was initially higher on play 3, and this reduction brings the groups' scores closer together on play 4.

There is a trend of increasing DE with each play of the game as expected, except for the final play for the reappraisal group, where a decrease due to cognitive reappraisal was expected. Scores were expected to increase as participants became more familiar with the simulation, when they could focus less attention on the game mechanics and more on their gains or losses while trading. Familiarity should allow participants to express their own trading behaviour more.

However, the reappraisal group has higher initial DE on play 3 than the control group. This is unexpected since the two groups were randomly assigned, and there was no difference in procedure between them until receiving the cognitive reappraisal instruction immediately prior to playing the simulation the final time. The experimental design was chosen to allow inferences to be drawn from equivalent groups, with the intention that the two groups had similar levels of DE prior to the reappraisal intervention being applied before play 4.

The difference between groups on play 3 and the effect of reappraisal seen on play 4 can be interpreted in two ways. One is that there was mean reversion between play 3 and play 4. Mean reversion occurs when differences at the extremes of a distribution occur by chance, and subsequent scores revert to the mean. If mean reversion was responsible then the reappraisal group just happened to have high DE scores on play 3, and the control group just happened to have low DE scores on play 3. Since these scores were both chance events, the differences unwind on play 4, leaving the means of the groups more likely to be equal.

An alternative explanation is that the reappraisal group had a higher intrinsic DE despite the random allocation of groups, and this was the reason for that group's higher mean on play 3. Without employing cognitive reappraisal, this group would have been expected to maintain their higher DE on play 4; however, the effect of reappraisal results in a decrease for this group on play 4.

6.4.1.1 Sources of between-participant variation in DE

These two explanations involve different sources of variation in DE scores between participants. Mean reversion invokes non-systematic noise, where each DE score observed differs from every other score due to random variation which is added to any “real” underlying score. The random component of each score will change every time the game is played, so will not persist between participants over time.

If these random differences happen to align with group membership on one play, such that one group happened to have mostly low scores while the other group had mostly high scores, then there could be a difference between groups. However, on the next play the expected difference between groups would still be zero, since these non-systematic differences change every time a measurement is taken. This explanation proposes that it is within-participant noise (the random variation of each participant’s DE each time it is measured), that is responsible for the group differences observed in DE, because of a chance alignment of within-participant noise with group membership on play 3.

The explanation from intrinsic differences invokes systematic variation between participants. Participants are assumed to have an underlying DE, which has a distribution across the sample of participants. For the difference between groups on play 3 to emerge, the participants randomly assigned to the reappraisal would on average have a higher intrinsic DE than the control group. Since this is a systematic difference, it would be expected to persist on play 4, before the effect of any reappraisal intervention is considered. This explanation proposes that it is between-participant differences (the intrinsic differences between participants of their disposition effects), that are responsible for the group differences observed in DE on play 3, because of a chance alignment of between-participant differences with group membership on play 3.

The latter explanation based on intrinsic difference is supported by two pieces of evidence. The first is that in the previous two chapters, considerable evidence was analysed for the claim that the disposition effect is a stable bias which differs between

people. The second uses evidence from scores on play 2. If differences between participants are intrinsic, then those differences would be expected to appear on play 2 as well (though there may be more noise due to participants being unfamiliar with the game, as discussed previously). As expected the reappraisal group is higher than the control group on play 2 too, albeit by a smaller amount than on play 3. To statistically analyse the effect of cognitive reappraisal using play 2, the next section reports the results of a marginal model using all three plays.

6.4.2 Marginal model of plays 2-4

A marginal model is used to test cognitive reappraisal while using scores from all three plays. Unlike a general linear model, this allows participants with missing data to be included in the analysis and is more robust to unequal sizes of groups. It also allows a covariance structure to be chosen that best fits the data, giving more reliable results from the statistical tests performed with the model.

Analysis using the marginal model has three stages. First, an appropriate covariance structure is selected. Then the fixed effects of the model are tested. A significant interaction between group and play order was predicted here, representing the effect of cognitive reappraisal. Finally, the individual betas in the model are tested. The interaction between group and play 4 was expected to have a negative beta, showing a decrease in the reappraisal group versus the control group.

6.4.2.1 Selection of covariance structure

A covariance structure was chosen by comparing the fit of each one, using Schwarz's Bayesian criterion (BIC), and the chi-square test between models (based on differences in deviance (-2LL), also known as the log-likelihood test).

A lower BIC indicates better fit. The testing of differences in deviance is more complex. Deviance is a measure of absolute fit of a model, and so deviance will always decrease when more parameters are added to a model. However, adding more parameters to a model may lead to data-fitting, fitting parameters to noise rather than a meaningful

underlying structure. The chi-square test for change in deviance judges whether additional parameters make a model significantly better fitting than a simpler alternative.

The difference in deviance between two models follows a chi-square distribution, with degrees of freedoms equal to the number of additional parameters added to the more complex model. A significant chi-square test shows that the additional parameters produce a decrease in deviance (indicating a better fit) which is significantly greater than would be expected.

Table 6.6 compares a variety of covariance structures using BIC, deviance, and the chi-square tests, comparing each one to compound symmetry. Every model makes a significant improvement on compound symmetry, based on the chi-square test. Compound symmetry makes two assumptions: the same variance for all repeated measures (in this model, the repeated measures are DE on plays 2-4) and constant correlation between all repeated measures. So, it appears that this combination of assumptions is not appropriate.

Table 6.6 Testing covariance structures for marginal model of DE scores

Covariance structure	Number of covariance parameters	BIC	Deviance (-2LL)	Deviance reduction vs CS	df	Chi-square p-value
Compound symmetry	2	213.3	203.3	n/a	n/a	n/a
Heterogeneous compound symmetry	4	212.2	192.1	11.2	2	.0037**
Huynh-Feldt	4	211.7	191.7	11.6	2	.0030**
Heterogeneous Autoregressive	4	216.4	196.3	7	2	.0302*
Heterogeneous Toeplitz	5	216.3	191.3	12	3	.0074**
Unstructured	6	220.0	189.9	13.4	4	.0095**

The latter three covariance structures reduce deviance significantly, but they also result in increases in BIC compared with compound symmetry. This shows that the models fit better, but that the improvement in fit is probably achieved by overfitting additional

parameters to the data. This leaves heterogeneous compound symmetry and Huynh-Feldt. They perform well, reducing BIC while also achieving the strongest results in reducing deviance, measured by the chi-square test p-values.

Both these structures allow the variances of repeated measures to vary independently. Heterogeneous compound symmetry assumes constant correlation between repeated measures, while Huynh-Feldt assumes the covariance between measures is equal to the average variance of each pair of measures, minus a constant. There are reliable correlations between repeated plays of the game, as seen in chapter 4, so heterogeneous compound symmetry is more intuitive in this situation. Since there is little difference in BIC between the two models, heterogeneous compound symmetry is selected for the covariance structure.

6.4.2.2 Tests of fixed effects

After selecting a covariance structure, fixed effects are added for play order, group, and the interaction between them. The main prediction is that there will be an interaction between play order and group: on play 4 the reappraisal group should display a significant decrease in DE relative to the control group. Results are shown in table 6.7.

Table 6.7 Testing fixed effects in marginal model of DE scores

	Model df	Residual df	F	p-value
Play order	2	95.7	4.92	.002**
Group	1	52.4	2.21	.143
Play order * group	2	95.7	3.29	.042*

As expected there is a significant interaction between play order and group, consistent with an effect of cognitive reappraisal. There is also a significant effect of play order, but no significant effect of group. These can be interpreted using figure 6.1 above, which plots DE scores: the effect of play order is that scores increase with each play; the interaction between group and play order is that the reappraisal group's DE average decreases on play 4.

Unlike an ANOVA, marginal modelling also allows estimation and testing of the individual betas used to represent the fixed effects. Play 2 of the control group is used as the baseline, and there are five betas in total for the model: two for play order (one each for plays 3 and 4), one for reappraisal group membership, and two for the interaction between play order and reappraisal group (again, one each for plays 3 and 4). The null hypothesis is that the beta is zero (i.e. has no effect), tested using one-sample t-tests; the results are shown in table 6.8.

Table 6.8 Testing individual betas in marginal model of DE scores

effect	beta	St error	t	df	p
Play 3	.143	.074	1.91	68.4	.060
Play 4	.219	.068	3.25	75.4	.002**
Reappraisal group	.280	.152	1.84	50.8	.072
Play 3 * reappraisal group	.113	.130	.870	71.4	.387
Play 4 * reappraisal group	-.209	.118	-1.77	77.8	.081

The tests of the betas support the interpretation of the fixed effects made earlier using figure 6.1. DE increases significantly over successive plays. Both betas for play 3 and play 4 are positive, indicating an increase in score from play 2. The beta for play 3 is marginally significant, while the beta for play 4 is significant. The beta for group is also positive, indicating a higher score for the reappraisal group, though it is only marginally significant.

The most important test though, is whether the beta for the interaction of play 4 and reappraisal group is significantly less than zero. (This represents the change in the reappraisal group on play 4 compared with the control group, once the reappraisal instruction has been given.) This beta is negative indicating lower scores for the reappraisal group as predicted. Although this effect is marginal, if this is treated as a one-tailed test (since the direction of the beta was specifically predicted in advance) then

it is significant, $p=.040$. The beta for the interaction between play 3 and group is positive indicating higher scores for the reappraisal group on play 3, but this is not significant.

This marginal model supports the original conclusion which tested the change in DE between plays 3 and 4. The differences in DE on play 3 are better explained as intrinsic differences, rather than mean reversion. The main driver of the overall interaction between group and play order is from the reappraisal group being lower on play 4.

If mean reversion was responsible for this effect, then including data from play 2 in the analysis would be expected to show that play 3 is the anomaly. There would be no significant interaction for the reappraisal group on play 4. In contrast, the interaction between reappraisal group and play order would be driven by the reappraisal group being higher on play 3, represented by a significant and positive beta for this effect. In these tests it is positive, but not significant.

6.4.2.2.1 Interpreting changes in DE

The estimated beta for the interaction between reappraisal and play 4 is $-.209$. This term can be exponentiated to produce a multiplier representing the effect of reappraisal on DE, as was seen earlier in this chapter. Exponentiated this equals 0.62 , which equates to a reduction of 38%. So, the marginal model shows that the average DE in the reappraisal group was reduced by 38% compared with the control group.

This effect is lower than the relative 55% reduction calculated earlier, from the change between play 3 and play 4. This is because some of the decrease in DE between play 3 and play 4 has now been accounted for by an increase in DE from play 2 to play 3. However, the reduction of 38% still represents an important reduction that can be achieved by using reappraisal.

6.5 CONCLUSION

The Milan study supports all three hypotheses set out at the beginning of this chapter. It was predicted that there would be a disposition overall, and this was confirmed. In addition, the disposition effect, tested separately on plays 3 and 4, was significantly greater than zero. Cognitive reappraisal significantly reduced the disposition effect as expected, and reappraisal resulted in a relative reduction of 55% in the disposition effect.

This reduction was expected to be driven by an increase in PLR, showing an increased willingness to sell losses: this was also supported. At the same time, PGR was not significantly different between groups. The experimental demonstration that the effect of reappraisal is due to its effect on holding losses is a novel finding.

There are some statistical caveats which limit the confidence which can be placed in the results of the study. A difference between the two groups in baseline scores suggested mean reversion might be driving the findings. This was addressed by analysing data using DE from plays 2-4 using marginal modelling. The evidence in favour of cognitive reappraisal's effect was not as strong, but still supported the predictions made.

The Milan study set out to test whether cognitive reappraisal could reduce the disposition effect while improving the external validity of the study. These results support that it does, verifying previous studies and showing that reappraisal's effect is not negated by more realistic testing environments.

7 THE EFFECT OF REAPPRAISAL ON THE DISPOSITION EFFECT AND EMOTIONS DURING TRADING IN NOVICES

The previous chapter presented the Milan study, which related to the third research question. It tested the effect of cognitive reappraisal on the disposition effect, cutting gains, and holding losses. This was done while also increasing the external validity of the test, by using a more realistic instrument to measure trading decisions, and by using a sample of retail investors as participants.

This chapter presents the OU study which relates to the fourth research question. It is similar to the Milan study since it also tests hypotheses relating to the effect of cognitive reappraisal on the disposition effect. Indeed, part of its aim is to build on the evidence about the effect of reappraisal from the Milan study. To this end, the first three hypotheses tested are the same as the Milan study:

- Novices will show a disposition effect
- Cognitive reappraisal will reduce the disposition effect
- Cognitive reappraisal will reduce holding losses but not affect cutting gains

Practical improvements were made to the Milan study, which allowed its previous findings to be tested more robustly. For example, this study features larger sample sizes, more manipulation checks, monetary incentives for good performance, and closer researcher control over participants during the study.

The cognitive reappraisal instruction may reduce emotions experienced by distancing the participant from their decision. This could reduce feelings of responsibility for decisions, and thus reducing the intensity of emotions during trading decisions, leading to less bias. So, reappraisal's effect on perceived responsibility is also included as a manipulation check of the reappraisal instruction.

However, the OU study does more than simply make practical improvements on the Milan study. This study is also intended to shed light on the mechanism of reappraisal's

effect on decision making. It was suspected that the effect of cognitive reappraisal in the Milan study was due to a reduction in emotions during trading, but the Milan study did not include variables to test this.

A mechanism involving emotions makes theoretical sense: cognitive reappraisal is a form of emotion regulation, and there is significant evidence that emotions (and particularly negative emotions) are involved in producing the disposition effect. So, the Milan also aimed to test two further hypotheses:

- Cognitive reappraisal will reduce negative emotions experienced during trading
- Changes in emotions during trading will mediate the effect of reappraisal

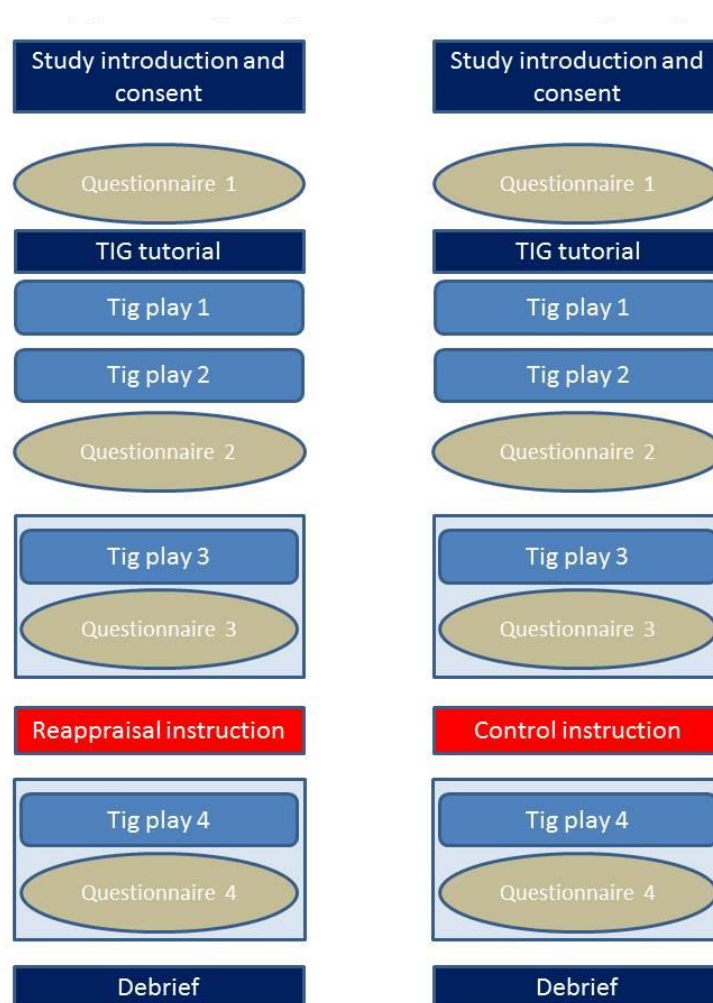
The chapter begins by describing methods used specifically in the OU study, including its experimental design, questionnaires used, and data management. It proceeds to discuss the results of the study in three parts. First, whether there is an overall disposition effect, then testing the effect of reappraisal on the trading biases and on emotions during trading. Finally, some exploratory analysis is conducted into why results were not as expected.

7.1 METHOD OF THE OU STUDY

7.1.1 Experimental design

The experimental design is similar to the Milan study, incorporating the same basic elements: 4 plays of the two-index game, plus questionnaires before and after some of the plays of the game. The effect of reappraisal is still tested in the same way, by comparing an experimental group to a control group, pre-intervention (play 3) and post-

Figure 7.1 Experimental design of the OU study



intervention (play 4). The OU study improved the manipulation checks of the reappraisal intervention, which were included in the questionnaires. Figure 7.1 shows an outline.

The design retains four repeated measures of the trading biases, despite time pressures from expanding the scope of the study. It was important to give participants sufficient time to become familiar with the game before reappraisal was tested. This was especially

the case since this study used novice participants. (Many participants struggled with the game initially and commented as such to the researcher. This was not unexpected, since many were very unfamiliar with the concepts of trading and stock markets. All participants confirmed that they understood the game before progressing to play 3.) Since the study was completed in person, the presence of the researcher allowed questions to be dealt with in person and further oral explanations were given to participants who found it difficult.

To create time for the emotion scale, while keeping the overall participation time around 1 hour, the duration of the two-index game was changed from 10 minutes to 5 minutes. In the Milan study, the number of options available to each participant increased throughout each 10-minute play (options include buying or short-selling, and trading in sizes of 1, 3, 5 or 10 shares). This format was only retained for play 1, and plays 2, 3 and 4 allowed all buying options from the beginning of the game. By allowing all options immediately a comparable trading record was produced for each participant in half the time.

A strength of the Milan study is that it used retail investors, rather than laymen or students. However, the practicalities of conducting data collection online, to work with retail investors, inevitably made it difficult to have tight control over the data. Many participants recruited in Milan did not complete the study online, so sample sizes were lower than intended, resulting in more noise in the group scores and less certainty about the effect of reappraisal observed. Among those that did, there were issues with non-compliance with the experimental protocol, which both lowered sample sizes further and unbalanced groups.

The OU study was completed in person, in temporary computer labs at the Open University. This achieved several practical aims. Since all participants were supervised by the researcher, much greater researcher control was possible over operational aspects of the study, leading to reduced data loss. Participation in person also allowed the study

design to include several long questionnaires (which if included in the online Milan study may have exacerbated dropout even more). The compromise for achieving this greater control was that the study was not carried out on a sample of retail investors, but on an adult sample drawn from staff and postgraduate students at the Open University.

However, this is considered some improvement on student samples, since it represents a broader range of the UK adult population.

Participants were given a paper booklet on arrival containing instructions and all the questionnaires. Appendix 4 is a full copy of this booklet for reappraisal participants (the control group booklet omitted sections relating to reappraisal). After a short introduction to the study and giving consent, participants completed questionnaire 1, which included the PANAS-30 scale. They were instructed to indicate how they felt "right now", to establish their baseline emotional state.

Participants completed a video-based tutorial of how to play the game and read written supplementary information. They could ask the researcher if they had any questions. Following this they played the game for the first time (play 1), and then a second time (play 2). Again, they could ask questions if they found the game difficult to understand. When they were satisfied that they understood the game, they moved to questionnaire 2, which included the resisting sunk costs scale (not reported, as the scale did not produce a consistent factor structure).

The remainder of the study produced the main experimental data. Participants completed play 3, followed by questionnaire 3 which included the PANAS-30 and additional emotions questions. Participants were instructed to indicate how they felt "when they were making decisions about whether or not to close positions".

Questionnaire 3 ended with the reappraisal instruction (or a control instruction for the group control), followed by play 4. Questionnaire 4 repeated the PANAS-30 and other emotion items, before moving to the manipulation checks (for the reappraisal group) and finishing with a few questions about game strategy, financial literacy and demographics.

After completion of the study, participants were debriefed on the disposition effect and the study's aims, given performance feedback, and informed of the results of the lottery for vouchers.

7.1.2 Reappraisal instruction

The full instruction can be found in appendix 4, in the instructions relating to the 4th play of the game. However, the main part of the reappraisal instruction is reproduced below:

When playing the two-index game this time, please imagine you are trading for someone else. You could be an investment manager trading on behalf of a client, or a pension manager, or simply someone managing investments on behalf of a friend or family member. Try to imagine this vividly when you are making decisions.

The control group booklet skipped this instruction, and simply reminded participants that their performance would determine their chances of winning a voucher (the reappraisal group were given the same reminder before the reappraisal instruction).

This instruction is similar to the Milan study, where participants were asked to imagine they were "in the role of an investment manager who is trading on behalf of a client". However, OU participants also have the option to trade on behalf of a friend or family member. The former will be referred to as "professional reappraisal" and the latter as "social reappraisal", and both are referred to as "reappraisal targets".

The social reappraisal option was included to help novice participants imagine trading for someone else. Novice participants may have little knowledge about what an investment or pension manager does, and may struggle to imagine how professionals would make trading decisions. This was less of a concern for retail investors, who were expected to have much higher financial literacy.

Reappraisal is suspected to reduce bias by allowing participants to emotionally detach from their decisions, and look at them more objectively. So, whether trading for a client

of acquaintance, what was thought to matter is that they were not trading for themselves. It was it's better to give novices a task they will be able to engage in as asking them to imagine something they had no experience of could invalidate the experimental intervention.

7.1.2.1 Manipulation checks

The OU study contained manipulation checks and questions about how participants employed reappraisal. Participants were also asked about their reappraisal target in questionnaire 4. The manipulation from the Milan study was also repeated; immediately below the instruction there was the following request:

Please tick the box to confirm that you have read the instruction in the paragraph above, about improving your performance:

Immediately before play 4 they received the following reminder:

Remember to imagine trading on behalf of someone else, though your aim is still to try to make as much money as possible. Good luck!

Miu and Crisan (2011) found that success in carrying out reappraisal was associated with changes in emotions experienced during reappraisal. To capture this, questionnaire 4 included some questions about the reappraisal process. Participants were asked how easy they found it to imagine trading for someone else. Difficulty of reappraisal was measured using a Likert item.

7.1.3 Questionnaires

Although the Milan study found an effect of reappraisal, it was not able to directly test hypotheses about the role of emotions in trading decisions. To address this, the OU study incorporates several emotion scales, duplicated both before and after reappraisal. The items aimed to measure emotion constructs which may be related to how the disposition effect is affected by reappraisal.

An ideal design would measure emotions at the point decisions are made. However, in the continuous decision-making environment of the two-index game, this wasn't possible. Participants made many decisions throughout the duration of the game, and disrupting this would vitiate the increase in ecological validity with the two-index game is intended to produce.

Participants were reminded immediately before answering the emotion scales that their responses should relate to how they felt when they were making decisions about whether to close positions or not. Of course, there are practical concerns with this approach. There will be memory issues, and participants should give responses which summarise their feelings over the many decisions which they have taken during each play of the game.

However, this is the most practical way to administer a questionnaire while also retaining features of the two-index which improve the ecological validity of the study. To make the game realistic, such as making multiple decisions in each play, making decisions in real time and making decisions in response to continuously changing information, the game cannot be stopped after each decision answer questionnaires. However, in support of this approach, Miu and Crisan (2011) gave participants the PANAS-X scale (discussed below) following a reappraisal instruction, and were still able to successfully measure changes in emotions.

7.1.3.1 PANAS items

The starting point for the questionnaire survey was the PANAS (Watson *et al.*, 1988). This comprises 20 items measuring positive and negative affect, listed in table 7.1. This is widely used scale in emotion research. For example, Hafenbrack *et al.* (2014) used the PANAS to study effects of emotion regulation (specifically mindfulness) on emotions in a study of the sunk cost bias. All 20 items from the PANAS were included in questionnaires 1, 3 and 4. The exact presentation and instructions for the PANAS can be seen in appendix 4.

Table 7.1 Positive and negative items in the PANAS

Positive affect items	Negative affect items
active	guilty
enthusiastic	afraid
attentive	nervous
excited	distressed
determined	hostile
strong	jittery
proud	irritable
alert	upset
interested	ashamed
inspired	scared

7.1.3.2 PANAS-X items

The positive and negative affect dimensions in PANAS reflect the valence of affect. However, emotions driving the disposition effect may be more specific than this, so measuring a wider range of emotional states is desirable (Summers and Duxbury (2012) made similar arguments, and are discussed in the next section). The PANAS-X (Watson & Clark, 1999) adds 40 additional items to the original 20 PANAS items to create “specific affect scales”. These subscales are intended to capture the content of emotional states, to supplement the two main dimensions which capture the valence of emotional states. There are 11 subscales which include negative emotions (fear, hostility, guilt, and sadness), positive emotions (joviality, self-assurance, attentiveness) and other affective states (shyness, fatigue, serenity, surprise).

Including all 11 would require all 60 items of the PANAS-X, tripling the length of the questionnaire. This might test participants’ patience, since the scale is given three times, and would considerably increase the duration of the study. However, as Watson & Clark suggest, a questionnaire can be kept to a reasonable size by including a selection of PANAS-X subscales. Since the original 20 PANAS items were already being used, some subscales from the PANAS-X could be included without adding many additional scale items.

Fear and hostility are two negative subscales in the PANAS-X which are most relevant to trading, the disposition effect, and reappraisal. Fear is associated with withdrawal or avoidance, so may promote avoidance of selling losses. Hostility on other hand is associated with approach behaviours; it is not expected to correlate with the DE and was included as a contrast to the effect of fear. Regret is the emotion most often linked the disposition effect, but the PANAS-X does not include subscale for this; regret is discussed in the section below.

There were no specific hypotheses about how positive affect and emotions could affect the disposition effect. However, attentiveness and serenity are useful emotions that could allow investors to make considered and rational decisions. They are a good contrast to the anxious and negative thinking that can result in poor trading decisions, which might reappraisal be expected to lessen, so they were also included.

Adding these subscales required an additional 10 items making a total of 30, more manageable than the full 60 items. This is referred to this as the "PANAS-30", and was included in questionnaires 1, 3 and 4. The items and subscales included are shown in table 7.2.

Table 7.2 Selected items from the PANAS-X used in the OU study

Fear	hostility	attentiveness	serenity
Afraid	disgusted	attentive	calm
nervous	hostile	determined	relaxed
shaky	scornful	alert	at ease
frightened	irritable	concentrating	
jittery	loathing		
scared	angry		

7.1.3.3 Additional emotion items

Rather than the PANAS or another emotion scale, Summers and Duxbury (2012) include different measures of emotions which they relate to "specific, task-related emotions" relevant to the disposition effect. They argue following Zeelenberg *et al.* (1998a) that merely experiencing a gain (loss) results in feeling elation (disappointment). However,

experiencing a gain (loss) while also feeling responsible for it results in rejoicing (regret). These are associated with specific action-tendencies (Zeelenberg *et al.*, 1998b). A disposition effect should be produced when someone feels responsible, and thus feels regret and rejoicing, but not when they do not feel responsible, and only experience elation and disappointment.

To test whether these facets of participants' emotion experiences were also affected by reappraisal, the OU study measured these specific emotions by adapting the items from Summers and Duxbury. Summers and Duxbury use Likert items to measure disappointment and regret, taken directly from Zeelenberg *et al.* (1998a). These items were also included in the OU study. However, their specific meanings of "elation" and "rejoicing" do not correspond to the normal meaning of these words, nor are there any other words that match their intended meaning.

Elation and rejoicing were measured by asking participants about their satisfaction with their trading outcomes (representing elation) and trading decisions (representing rejoicing). Unfortunately, this method was not possible in the OU study. To attempt to measure similar positive emotions, participants were asked how happy, satisfied and proud they felt.

The reason the original items could not be included was that the design of Summers and Duxbury (2012) isolated one decision at a time for participants, allowing participants to respond about their feelings when making one specific decision. In the OU study participants make many decisions over a 5-minute period, so asking about single decisions was not possible.

The distinction between outcomes of decisions and responsibility for decisions was also difficult to incorporate into the OU study. As discussed above, participants are given an instruction immediately before completing the emotion questionnaire that they should answer with reference to how they felt when they were contemplating selling during the

two-index game. So, asking participants to respond about how they felt regarding the outcome of the decisions could easily confuse participants.

Summers and Duxbury also asked participants how responsible they felt for the outcomes of their decisions. This was a manipulation check on experimental design, which attempted to manipulate feelings of responsibility (and in turn affect the disposition effect). So, responsibility was also included in the OU study as a manipulation check, to test whether cognitive reappraisal does affect responsibility, which in turn affects the disposition effect.

7.1.4 Data management

Like the Milan study, the two-index game produced measurements of DE, PGR and PLR. Each participant has scale scores for: positive affect (PA), negative affect (NA), fear, hostility, attentiveness, and serenity, each measured in questionnaires 1, 3 and 4. The items based on Summers & Duxbury are single item variables. "Change scores" were created for all these variables, using the difference between play 3 and play 4.

7.1.4.1 Missing data

There were many variables and participants missing at least one data point, largely because of the large number of items required to be answered over the 4 questionnaires. ("Variable" here refers to each piece of information supplied by participants, so each item of each questionnaire is a variable). Figure 7.2 and table 7.3 provide a breakdown of missing values.

The first pie chart in figure 7.2 shows that two thirds of variables had at least one data point missing, and table 7.3 gives a more detailed breakdown of how many missing data points there were across all variables. The second pie chart in figure 7.2 shows that slightly less than 50% of participants ("cases") were missing at least one data point. However, overall there was less than 3% of data missing, shown in the third pie chart, so the information available in the dataset was high even though the number of complete variables and cases was low.

Figure 7.2 Summary of missing values

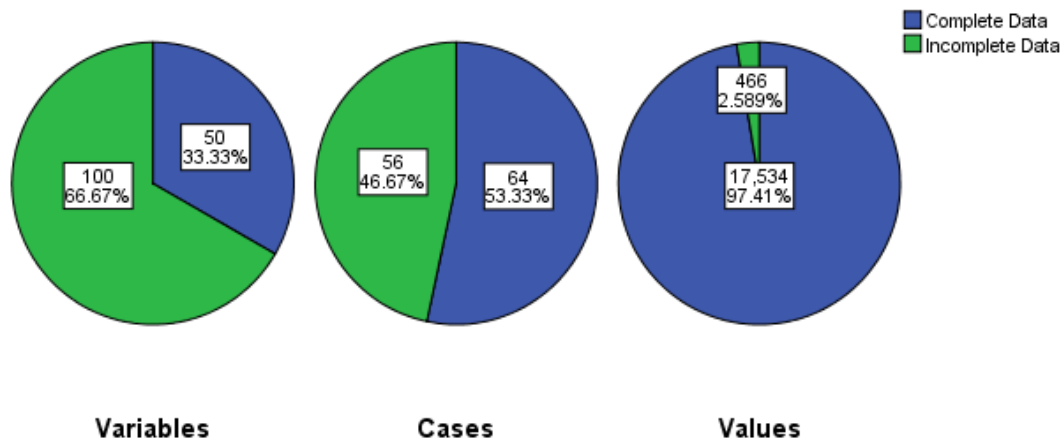


Table 7.3 Number of missing values per variable

Number of missing values	Number of variables	Cumulative variables	Cumulative % variables
0	50	50	33
1	40	90	60
2, 3 or 4	30	120	80
5 or 6	7	127	85
10-14	16	143	95
15-19	7	150	100

7.1.4.2 Multiple imputation

To create scale scores from the many individual items, complete data is needed. However, removing cases with listwise deletion would result in dropping many participants, despite the low amounts of missing data in the sample. To solve this, missing values were replaced using multiple imputation. (The process of multiple imputation was described in chapter 5, when it was applied to the disposition effect scale in the Milan and London studies.)

In the OU study, the number of individual questionnaire items meant many variables needed to be imputed; in fact, there are more variables than participants. This can cause problems with multiple imputation, so to use multiple imputation within its limitations, scale scores from the questionnaires were imputed rather than the individual scale items. This drastically cut the number of variables imputed. The items based on

Summers & Duxbury are single item variables so they needed to be imputed as single items. Trading bias scores from the 4 plays of the two-index game were also imputed, although there little missing data for these.

Multiple imputation also allows other variables to be included as predictor variables. These are not imputed themselves, simplifying the mathematical process, but can be used to help identify relationships for predicting the imputed variables. Predictor variables used included: all the individual PANAS-30 items (30 items each from questionnaires 1, 3 and 4); group membership; and the difficulty of reappraisal rating for the reappraisal group. 10 imputations were created, and the results below are pooled across all 10 imputations, unless noted otherwise.

7.2 OVERALL DISPOSITION EFFECT

An overall disposition effect for the two groups combined was observed on each of the four plays, and also across the entire study (combining data from the four plays). As figure 7.3 shows, the mean disposition was significantly greater than zero (and since DE has been logged, a DE of 0 after exponentiation corresponds to a PGR/PLR ratio of 1, i.e. losses are as likely to be sold as gains). The plays were not significantly different from each other, as seen in the large overlaps in confidence intervals.

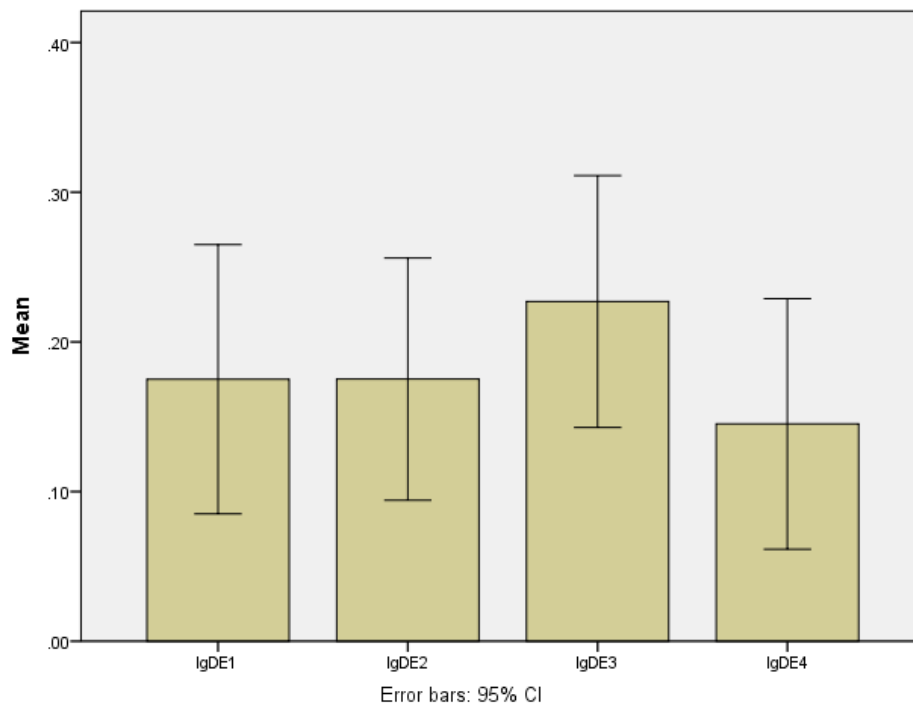
Figure 7.3 Disposition effect on each play

Table 7.4 details descriptive statistics for the means DE scores, and the PGR/PLR ratio from exponentiating DE. This ranges from 1.18 to 1.25, with an average of 1.21, meaning that gains were about 20% more likely to be sold than losses. The size of this effect is much lower than the Milan study, and also other research on the disposition effect; however, it is still a significant effect. So, participants were (statistically) significantly more likely to sell losses, but not by as much as usually seen.

Table 7.4 Disposition effect on each play

	N	Mean DE	Std. Error	PGR/PLR ratio
play 1	120	0.192	0.058	1.21
play 2	120	0.191	0.043	1.21
play 3	120	0.217	0.042	1.24
play 4	120	0.155	0.043	1.17
All plays		0.189		1.21

7.3 THE EFFECT OF COGNITIVE REAPPRAISAL

7.3.1 Manipulation checks

7.3.1.1 *Target of reappraisal and difficulty of reappraisal*

All participants in the reappraisal group checked the manipulation check box to confirm they had read the reappraisal instruction. Unfortunately, some responses about their reappraisal target cast doubt on whether they had carried out reappraisal properly, so they were excluded from the analysis. There was also one participant who did not make any trades on plays 3 and 4, so was excluded from analysis.¹⁴

In addition to checking that reappraisal had been carried out, the manipulation checks also measured how easily it was carried out. Difficulty was measured out of 5 using a self-report Likert item. Participants answering either 4 (hard) or 5 (very hard) were excluded from analyses which tested reappraisal, since they were at risk of not having carried out reappraisal effectively. These comprise most participants who were excluded from the reappraisal group.

Table 7.5 details the effect of these manipulation checks on the number of participants included in the study. The randomisation of participants between groups was weighted slightly to reappraisal in anticipation of data attrition from that group. However, it did not completely mitigate this, and the control group ended up slightly larger.

¹⁴ This was actually an interesting case of “rationality” in decision making, and whether behaviour in experiments can be generalised to real life. After perceiving that they performed poorly on plays 1 and 2, the participant decided that their attempts at making profitable trades only made things worse, and their best chance of winning a voucher would be if they did make no trades and finished the game with zero profit.

Table 7.5 Participants excluded by group

Category	Control	Reappraisal
Included	51	40
Excluded	1	28
Did not use reappraisal	-	4
Dubious reappraisal target	-	3
Rated "very hard"		4
Rated "hard"		17
Made no trades	1	-
Original total	52	68

7.3.1.2 *The effect of reappraisal on responsibility*

The effect of reappraisal on perceived responsibility was included as a manipulation check. In the Milan study, the logic of the reappraisal intervention was that managing investments for someone else, rather than for oneself, would reduce perceived responsibility for positions. If responsibility was lower, then the intensity of emotions experienced was expected to decrease, and result in a lower bias in decision making. The reduction in the disposition effect in the Milan study supported this, but it was not tested directly.

The effect of reappraisal was tested in the same way as in the Milan study, using the change in score from play 3 to play 4, comparing the reappraisal group and control group. Table 7.6 shows the absolute change by group. Contrary to expectations, the reappraisal group increases in responsibility, and this difference is significant as reported in table 7.7.

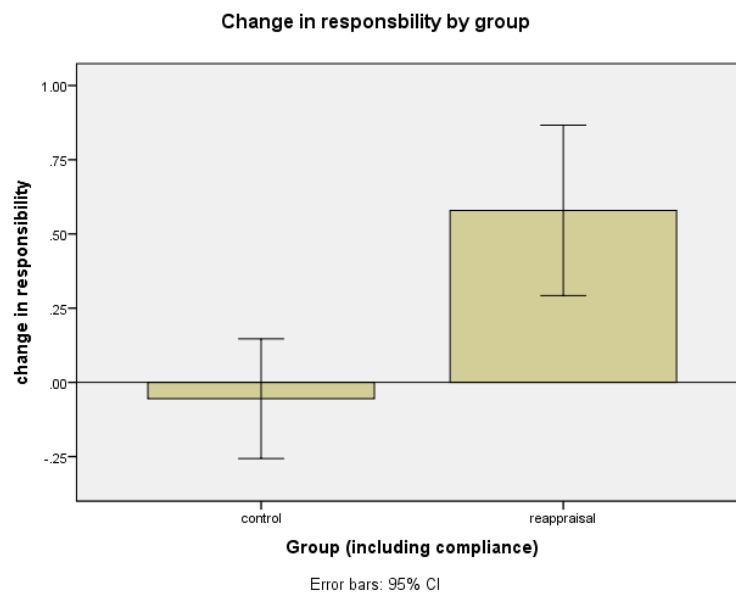
Table 7.6 Change in perceived responsibility by group

	Group	N	Mean change	Std. Error
responsible responsible	Control	51	-.05	.10
	Reappraisal	40	.56	.15

Table 7.7 T-test of change in perceived responsibility between groups

	Mean difference	Std. error	t	df	p-value
responsible	.61	.19	3.33 ¹⁵	331	.001**

The relative change was 0.61 points, which is also of practical importance on a scale which only has a range of 4 points (i.e. between 1 and 5). The changes analysed separately by group tell the same story: the control group change is not significantly different from zero, while the reappraisal group's increase of 0.56 is significant. Figure 7.4 illustrates this difference between groups.

Figure 7.4 Change in responsibility by group

7.3.2 The effect of reappraisal on DE, PGR and PLR

The effect of reappraisal on the trading biases are tested in the same way as in the Milan study, using the change in score from play 3 to play 4, comparing the reappraisal group and control group. Reappraisal was expected to decrease DE, by increasing PLR while not affecting PGR.

¹⁵ Levene's test was significant, so this test did not assume equal variances

Table 7.6 shows the absolute changes by group. A brief look at these figures suggests that the hypotheses are not supported. Both groups decrease in DE, but the control group decreases by more than the reappraisal group. Both groups increase in PLR, but the control group increases by more. Changes in PGR for both groups are close to zero, so the null of reappraisal on this score is supported.

Table 7.8 Change in trading biases by group

Change in..	Group	N	Mean change	Std. Error
DE	reappraisal	40	-.0685	.0589
DE	control	51	-.0976	.0565
PGR	reappraisal	40	-.0246	.0507
PGR	control	51	-.0156	.0439
PLR	reappraisal	40	.0439	.0646
PLR	control	51	.0820	.0611

Independent t-tests are carried out for each bias to confirm these findings, using group membership as the independent variable. These results are summarised in table 7.7. The mean difference is calculated as the change in reappraisal group minus the change in the control group. All the differences between groups are small and none are significant, nor approach significance. Given that contrary to the hypotheses, reappraisal did not affect the trading biases, the mediation of reappraisal's effect by change in emotions is not tested later in this chapter.

Table 7.9 T-tests of change in trading biases between groups

p	Mean Difference	Std. error of difference	t	df	p-value
DE	.0290	.0811	.358	137	.721
PGR	-.0090	.0670	-.134	1902	.893
PLR	-.0380	.0886	-.429	689	.668

7.3.3 The effect reappraisal on emotions

The same method is used to test changes in emotions during trading, comparing scores on play 3 with score on play 4. Reappraisal was predicted to reduce negative emotions during trading. Positive emotions were not expected to be affected, but were also tested to rule out the possibility that changes in these emotions could explain reappraisal's

effect on trading biases. Table 7.8 shows the absolute changes by group, and table 7.9 shows independent t-tests that test the difference between groups (Levene's test was significant, so neither test assumes equal variances).

Both groups show a small reduction for both positive and negative affect, but there were no significant differences between groups. This suggests that both groups simply became slightly less emotional when they played the game again. The changes in scores were significant for the control group analysed separately. However, as each scale has a theoretical range of 10 to 50, the mean changes found are very small in practical terms.

Table 7.10 Change in positive and negative affect by group

	Group	N	Mean change	Std. Error
Positive affect	Control	51	-1.19	.51
Positive affect	Reappraisal	40	-1.58	1.03
Negative affect	Control	51	-1.40	.44
Negative affect	Reappraisal	40	-.539	.76

Table 7.11 T-tests of change in positive and negative affect between groups

	Mean Difference	Std. error	t	df	p-value
Positive affect	-.387	1.15	-.337	396766	.736
Negative affect	.861	.834	1.03	34484	.302

Table 7.10 shows the absolute changes of the PANAS-X scales split by group. The only generalization that can be made might be it that all scores (except serenity in the control group) tend to decline by a small amount from play 3 to play 4. However, this decrease is seen in both groups¹⁶ so it is not an effect of reappraisal, but of playing the game or responding to questionnaires repeatedly.

¹⁶ This absolute change, analysed in each group separately, was significant in the control group for the decreases in fear and attentiveness, and the increase for serenity. No changes were significant for the reappraisal group. However, no hypotheses were made about absolute changes in advance, and as noted the difference between the changes by group was not significant.

Independent t-tests confirm this, reported in table 7.11. There are no significant differences between groups, so no effect of reappraisal. Nor is there a clear pattern contrasting control and reappraisal groups. The control group decreases more in fear and attentiveness, but increases more in serenity. Hostility in both groups is almost the same. For all emotions though, the difference between groups are small compared with the possible range of scores: all differences between groups were smaller than 1 point on a single item. Fear and hostility could range from a score of 6 to 30, attentiveness from 4 to 20, and serenity from 3 to 15, so all relative differences were small in practical terms.

Table 7.12 Change in PANAS-X scales by group

	Group	N	Mean change	Std. Error
Fear	Control	51	-1.19	.33
	Reappraisal	40	-.48	.58
Hostility	Control	51	-.31	.36
	Reappraisal	40	-.42	.32
Attentiveness	Control	51	-.69	.24
	Reappraisal	40	-.26	.35
Serenity	Control	51	.58	.19
	Reappraisal	40	-.03	.45

Table 7.13 T-tests of change in PANAS-X scales between groups

	Mean difference	Std. error	T	df	p-value
Fear	.71	.63	1.13	9839	.260
Hostility	-.11	.50	-.22	6875	.826
Attentiveness	.43	.42	1.03	15073	.303
Serenity	-.61	.45	-1.33 ¹⁷	18598	.217

The additional emotion items are tested in the same way. Table 7.14 shows the absolute changes by group, and table 7.15 shows t-tests of the differences between groups. All changes are small and none of the differences are significant between groups.

¹⁷ Levene's test was significant, so this test did not assume equal variances

Table 7.14 Change in other emotion items by group

	Group	N	Mean change	Std. Error
regretful	Control	51	-.12	.16
	Reappraisal	40	-.38	.15
regretful disappointed	Control	51	-.39	.19
	Reappraisal	40	-.41	.20
happy	Control	51	-.08	.15
	Reappraisal	40	-.06	.17
satisfied	Control	51	.08	.15
	Reappraisal	40	-.01	.20
proud	Control	51	.14	.13
	Reappraisal	40	-.07	.19

Table 7.15 T-tests of change in other emotion items between groups

	Mean difference	Std. error	t	df	p-value
regretful	-.25	.23	-1.10	9527	.270
	-.02	.28	-.07	59766	.944
disappointed	.02	.24	.07	5556	.942
	-.09	.24	-.368	263998	.713
happy	-.22	.22	-.977	9127	.328

7.4 FOLLOW-UP ANALYSIS TO NULL RESULTS

Many of the results found were unexpected. Reappraisal did not affect any trading biases in the OU study, which directly contradicts the findings of the Milan study, and also the findings of Lee *et al.* (2008). Nor did reappraisal have any effect on any emotions except for perceived responsibility, which was the opposite effect to what that expected. To investigate this, some follow-up tests are reported using participants' rating of the difficulty of implementing reappraisal, and the reappraisal targets used by participants.

7.4.1 Effect of difficulty of reappraisal

In the tests above, the difficulty rating was only used as an exclusion criterion: participants who rated reappraisal as hard or very hard were excluded. In the tests that follow, all participants are included to try and establish how the effect of reappraisal may vary when participants find it easier or harder to implement reappraisal, and what effects reappraisal may have on decision making in general.

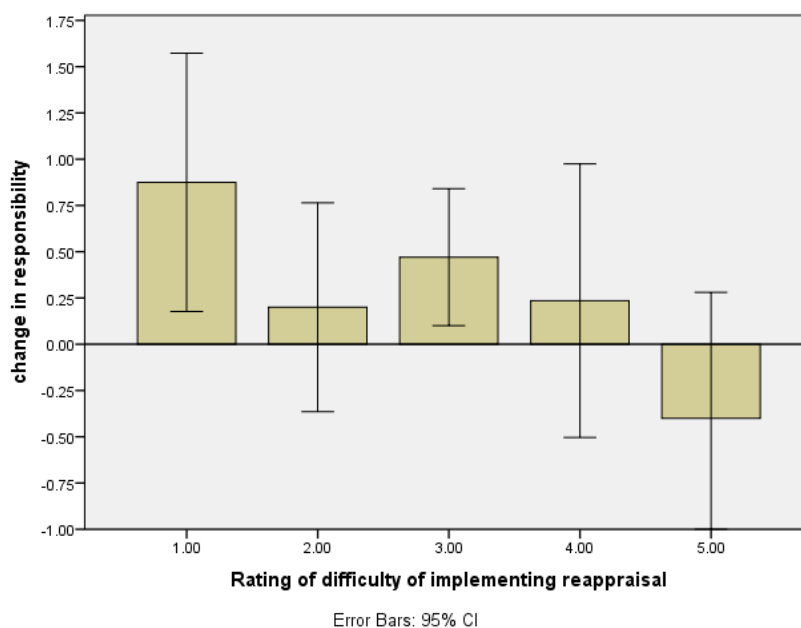
7.4.1.1 Difficulty of reappraisal and perceived responsibility

There is a marginal negative correlation between change in responsibility and the difficulty rating of reappraisal¹⁸ ($r = -.233$, $p = .061$). As table 7.14 and figure 7.5 show, the change is mostly apparent at the extremes of the range: participants who found reappraisal “very easy” had larger increases in perceived responsibility while those ratings it “very hard” actually decreased in perceived responsibility.

Table 7.16 Change in responsibility in the reappraisal group split by difficult rating

Difficulty rating	Change in responsibility	Std. error	N
1	.88	.30	8
2	.20	.25	10
3	.44	.21	28
4	.17	.32	17
5	-.40	.24	5

Figure 7.5 Change in responsibility in the reappraisal group split by difficulty rating



¹⁸ This is significant in the raw data ($r = -.289$, $p = .020$). In the imputed datasets, participants with difficulty ratings of 4 or 5 who had missing data for responsibility had comparatively high imputed scores for change in responsibility, moderating the correlation.

This result is more evidence that reappraisal increases perceived responsibility, and that the self-report difficulty measure is a reasonable proxy of how effectively each participant implemented reappraisal. The easier participants found it to implement reappraisal, the more likely their perceived responsibility was higher. Of course, this interpretation is still tentative: splitting the reappraisal group into five groups greatly reduces the statistical power achieved by this study, as can be seen by the wide confidence intervals in the chart.

7.4.1.2 Difficulty of reappraisal, and effect of reappraisal on the disposition effect

The lack of effect of reappraisal on the disposition effect was very surprising. In fact, the reappraisal group fared worse than the control group (though this difference was not significant). The next test looks at whether the effect of reappraisal was associated with participants' difficulty rating of reappraisal. Table 7.15 reports the changes in DE split by rating.

Table 7.17 Change in DE in the reappraisal group split by difficult rating

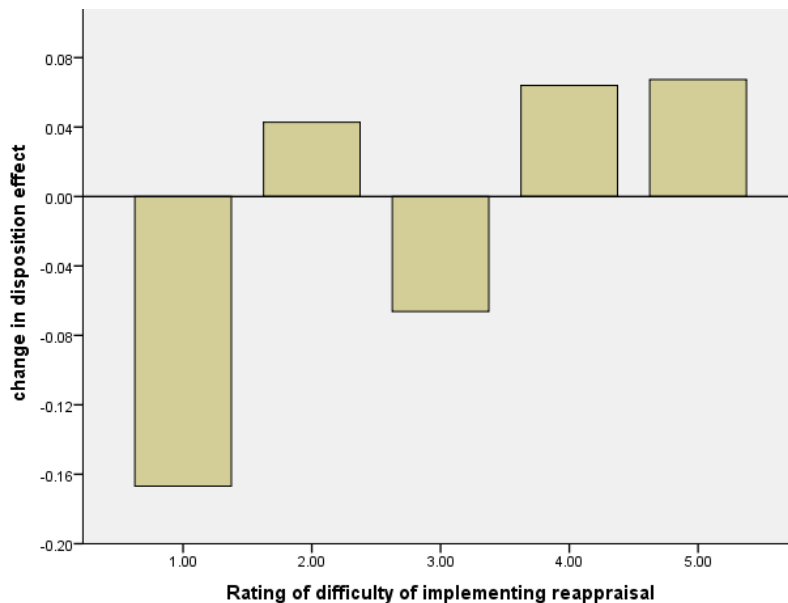
Difficulty rating	Change in DE	Std. error	N
1	-.17	.08	8
2	.02	.17	10
3	-.07	.08	28
4	.07	.13	17
5	.07	.13	5

There is a rough trend of increasing change in disposition effect from play 3 to play 4, as difficulty increases: participants who found it difficult to implement reappraisal tended to increase on play 4. However, this correlation is not significant ($r = .166$, $p = .201^{19}$). As seen in figure 7.6, the change is mostly apparent at the extremes of the range: participants who found reappraisal very easy had larger decreases in disposition effect

¹⁹ Though this is marginally significant in the raw data ($r = .210$, $p = .092$). In the imputed datasets, participants with difficulty ratings of 2 who had missing data for change in disposition effect had comparatively high changes imputed, attenuating the correlation.

than those who found it very hard.

Figure 7.6 Change in DE split by difficulty rating



Looking at the breakdown of DE into PGR and PLR, this trend is driven by a negative correlation between difficulty rating and change in PLR: increasing difficulty results in a decreased PLR representing an increased tendency to hold losses. For people who rated difficulty low, the opposite is true: they tended to have a larger increase in PLR, representing a decreased tendency to hold losses, and therefore a lower disposition effect. Although this is not significant either ($r = -.152$, $p = .244$), the correlation for PGR is effectively zero ($r = .015$, $p = .904$). So, to the extent that we give any credence to the trend between DE and difficulty, it is attributable to the relationship between PLR and difficulty.

As with the results of difficulty on perceived responsibility, the trend is only suggestive. The results are not significant, but this is not surprising given the very low power that results from splitting the reappraisal group into five groups. However, the results for participants rating reappraisal very easy are intriguing. This subset showed a large decrease in disposition effect (a decrease of $-.17$, which exponentiated represents a decrease of 32% in the PGR/PLR ratio). This is a greater decrease than the control group achieved, shown in the analysis earlier in this chapter (decrease of $.10$, an

exponentiated decrease of 20%). So, this is very tentative evidence that reappraisal may only be effective for people who find it easy to carry out.

In fact, the trend in figure 7.6 is evidence to support this view too, regardless of the absolute decrease in participants who found reappraisal very easy. Reappraisal had a more negative effect on the disposition effect of participants who found it difficult to implement, and the reduction in disposition effect from reappraisal improved as participants found it easier to perform. The association between change in disposition effect and difficulty rating suggests that the experimental intervention in this study may have been hindered by other aspects of implementing reappraisal.

A possible explanation for this is that the cognitive demands of implementing reappraisal interfered with decision making: participants who use more cognitive resources on implementing reappraisal had fewer available to play the game, and this increased the bias they displayed. It is notable that many participants found the game alone taxing, without any complications from simultaneously imagining they were taking decisions for some other party.

Combining the correlations of difficulty with changes in responsibility and the disposition effect, it appears that an increase in perceived responsibility does not explain the null effect of reappraisal on the disposition effect. Participants who found reappraisal easiest had the largest increase in perceived responsibility, but also saw the most benefit from implementing reappraisal (i.e. the largest drop in disposition effect). Again, these results are tentative because of the small samples, but the results which are available do not support increased responsibility as an explanation.

7.4.2 Reappraisal targets used

Another factor which may have influenced the effect of reappraisal is how reappraisal was implemented. As discussed earlier, OU participants had a choice of reappraisal target. A manipulation check in questionnaire 4 asked "who did you imagine you were trading on behalf of?" Participants fell mainly into professional reappraisal (i.e. as a

financial professional) or social reappraisal (for a friend or family member). These categories, together with the difficulty rating for each participant, are shown in table 7.16.

Table 7.18 Reappraisal targets and reappraisal difficulty rating

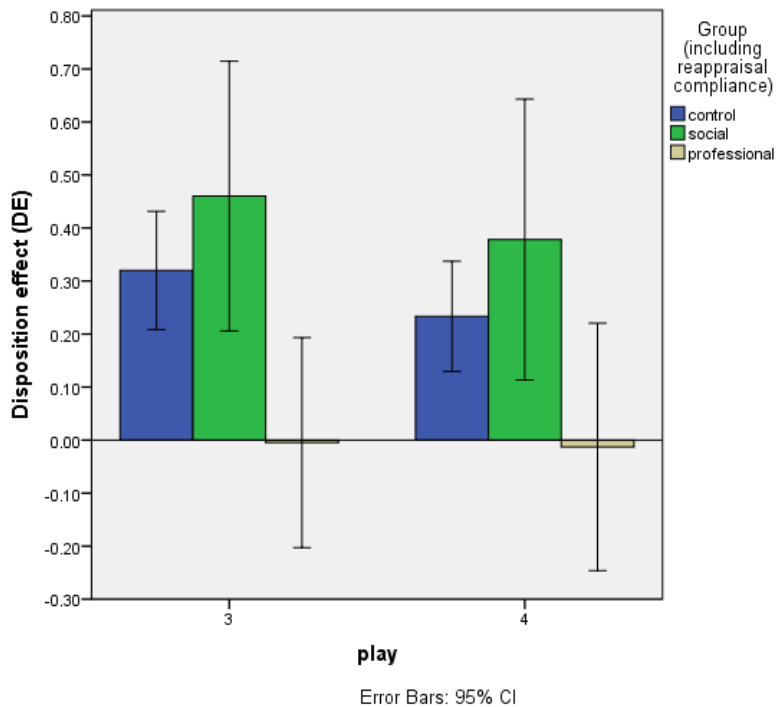
Type	Total	Difficulty rating 1, 2 or 3	Difficulty rating 4 or 5
Social reappraisal	25	16	9
Professional reappraisal	28	18	10
Anonymous reappraisal	8	6	2
Reappraisal – dubious	3	n/a	n/a
Reappraisal – blank	4	n/a	n/a
Control	51	n/a	n/a
Control – no trades	1	n/a	n/a

The anonymous reappraisal category is where participants did not give enough information in their answer to place them in one of the main categories. “Dubious” and “blank” are literal descriptions of those reappraisal participants’ responses about their reappraisal target, so they were also excluded from the analysis above. One control participant also made no trades, also as discussed above.

7.4.2.1 Analysis of reappraisal using reappraisal targets

To test if there were differences between groups, the disposition effect on plays 3 and 4 was re-analysed, comparing professional reappraisal, social reappraisal and the control group. Other reappraisal participants were excluded, as well as those with a 4 or 5 difficulty rating. Figure 7.7 presents these results.

Figure 7.7 DE by group and reappraisal target on plays 3 and 4



The disposition effect increases in the control and social reappraisal groups, and there is no significant difference between these two groups. The disposition effect for the social reappraisal group is a little higher than the control group, but this difference is present on both plays 3 and 4.

In contrast, the professional reappraisal group is significantly different from both the control and social reappraisal groups. Its overall disposition effect was not significantly different from zero on either play 3 or play 4. To put it another way, even before implementing reappraisal, the professional reappraisal group was just as likely to sell losses as sell gains. Therefore, the professional reappraisal group couldn't reduce their disposition effect, as on average they didn't have one to begin with.

This is certainly an unexpected finding. It seems that the kind of people who choose to implement reappraisal as a professional have a low disposition effect initially. However, it still doesn't provide a clear explanation why reappraisal is ineffective in the social reappraisal group, which is statistically the same as the control group. It does suggest

though that the exact application of cognitive reappraisal, and the choice of reappraisal target, may be more complex than anticipated.

7.5 CONCLUSION

The OU study was successful in its practical aims, collecting a much larger, balanced sample size. There was less missing data, and greater confidence that participants had complied with the protocol as a researcher was always present during participation.

There was a disposition effect overall; however, there was no effect of reappraisal on the disposition effect, nor cutting gains, nor holding losses. The only significant effect of reappraisal was the manipulation of changes in perceived responsibility for the outcome of their decisions, finding reappraisal produced an increase. The null effect of reappraisal on emotions is not surprising considering the null effect of reappraisal on the trading biases, which suggests that reappraisal was ineffective in general in this study.

Emotion measures were included to investigate if they mediate the effect of reappraisal. There was no effect of reappraisal on emotions, and no effect of reappraisal on trading behaviours to mediate, so no firm conclusions can be drawn about whether emotions mediate a reduction in DE when one does occur.

As noted, the only significant effect of reappraisal was an increase in perceived responsibility for the outcome of their decisions. However, this was the opposite of what was expected. Lee et al. (2008) found that a reappraisal instruction to trade for someone else decreased the disposition effect, and speculated that this was the result of enabling participants to distance themselves from their decisions. In turn, the hypothesis in this study was that an increased distance from decisions should decrease the emotions experienced when making those decisions, thus reducing trading biases. Instead, when trading for someone else, participants appear to feel *more* responsible than if they were trading for themselves. In retrospect, this reaction by participants is understandable, and was probably amplified by the option to trade for a friend or family member, rather than an anonymous client. This is discussed more in the limitations section of chapter 9.

Follow-up analysis suggests that differences some participants found it difficult to carry out reappraisal and that this also interfered with its expected effect. Perceived responsibility increased more in participants who rated reappraisal easier to carry out. This reinforces the conclusion above that reappraisal had the opposite effect on responsibility to the one intended. Note however, that this correlation was only marginal. In addition, the disposition effect decreased more in participants who rated reappraisal easier to carry out. This effect was strongest for those who rated reappraisal 'very easy', though the overall correlation is not significant. However, it does suggest that the increased responsibility from reappraisal was not responsible for reappraisal not being effective at reducing the disposition effect overall.

For participants who found reappraisal difficult, it is possible that playing the game while also attempting cognitive reappraisal was too cognitively demanding, and that the emotional impact of gains and losses was crowded out by the cognitive demands of simultaneously attempting the game and reappraisal. This would be more likely to occur with novices as participants, who are less familiar with trading tasks than retail investors. Overall the study suggests that this use of reappraisal in novices is not be straight-forward, and this is discussed in the last two chapters.

8 SUMMARY OF FINDINGS

The purpose of this chapter is to summarise the results of the empirical work in this thesis. It clearly sets out and discusses the evidence presented for each research question and their associated hypotheses, and whether each hypothesis was supported. This allows the contributions, implications and limitations to be discussed in the final chapter, without needing to refer to the tests in detail.

This thesis covers four research questions:

RQ 1 - Does the disposition effect have trait-like characteristics?

RQ 2 - Do cutting gains and holding losses have trait-like characteristics?

RQ 3 - Does cognitive reappraisal affect the disposition effect and its constituent biases, when tested in experienced traders under conditions of greater external validity?

RQ 4 - Does cognitive reappraisal affect the disposition effect and its constituent biases, by changing emotions during trading, when tested in novices under conditions of greater external validity?

These questions can be divided into two themes. The first two research questions investigate the trait-like characteristics of trading behaviour, while the second two research questions explore whether it is possible to change these trading behaviours and explore their links to emotions during trading.

The first research question looks at the properties and measurement of the disposition effect. The second research question looks at similar issues about its constituent biases (cutting gains and holding losses), with the additional issue of whether those constituent biases are independent, and thus the disposition effect is better conceived as the combination of these constituent biases. The third and fourth research questions examine whether it is possible to change the disposition effect people display when making trading decisions, by using cognitive reappraisal in ecologically realistic experimental conditions, in both experienced and novice participants.

The four sections below recap the hypotheses relating to each research question, how they were tested, whether the results support the hypothesis, and the conclusions drawn.

8.1 DOES THE DISPOSITION EFFECT HAVE TRAIT-LIKE CHARACTERISTICS?

This research question examines the disposition effect by looking at two important properties. First, is the disposition effect stable when measured: do the same participants achieve similar scores when measured repeatedly? Is it meaningful to say that a person has a high or low disposition effect, as if it was a stable personality trait? Second, does the disposition effect have construct validity: does it correlate across measures and situations? If the disposition effect is a stable behaviour that drives behaviour, it should have both intra-individual stability and validity.

If the disposition effect has trait-like characteristics, it should be possible to measure a person repeatedly and get roughly the same answer. If this is not the case, it raises questions about whether the disposition effect is better seen as a state phenomenon rather than a trait-like phenomenon. If this were the case, investors could differ in their disposition effects, but this would not be due to differences in an underlying disposition effect; it would not make sense to refer to investors as high or low in disposition effect, outside of the specific occasion where they were measured. This would nullify the rationale of identifying investors prone to the disposition effect, to help them reduce their bias.

One way to investigate construct validity is whether different ways of measuring the disposition effect give similar answers, and in particular whether measurements in the lab relate to disposition effects in the real world. The assumption here is that studying the disposition effect in an experimental setting is ultimately motivated by its occurrence in real trading. This is a phenomenon waiting for a definitive explanation: even Shefrin

and Statman's seminal 1985 paper sought to test explanations for its appearance in financial markets.

Table 8.1 shows results for the hypotheses relating to the first research question, and then each hypothesis is discussed briefly.

Table 8.1 Findings for trait-like characteristics of the disposition effect

	Hypothesis	Test	Supports hypothesis	Chapter
1.1	The disposition effect will show intra-individual stability	Correlations between repeated plays of the two-index game	Yes	4
		Amount of variance in disposition effect scores accounted for mainly by differences between participants	Yes	4
		Test-retest stability (in the Milan study)	Yes	4
1.2	The disposition effect will show convergent validity across multiple measures	Correlation between DE scores from the two-index game, and disposition effect scores from real-world trading records	Yes	4
		Correlations between DE scores from the two-index game, and disposition effect scale scores.	Yes	5

8.1.1 The disposition effect will show intra-individual stability

Participants played the two-index game repeatedly in the same experimental session (3 or 4 times depending on the study). Strong correlations were expected between disposition effect scores (DE scores) on repeated plays. This is strongly supported, with all correlations being significant with a typical correlation coefficient of 0.7. The evidence is particularly strong because this was done with 3 independent samples, and the same relationship holds in all 3 studies.

Another technique to measure the stability of scores partitioned the variance within each sample into within-participant and between-participant variance. If scores are expressions of a stable behaviour which differs from person to person, then most variance would be expected to be attributable to between-participant differences. Again,

this is what was found, with the between-participant variance varying between 60-75% depending on the sample.

Finally, if the disposition effect is a stable behaviour, scores from repeated measurements should correlate over extended periods of time. In the Milan study, participants had a break between plays of the game of a week or more; however, the test-retest correlations are still significant and strong ($0.568 < r < .732$). This last test suggests that the disposition effect is stable not just in the short term (the same hour), but also over the medium term (a week or two).

Based on these results, we can conclude that the two-index game produces reliable scores of the disposition effect in the short term, and probably in the medium term also.

8.1.2 The disposition effect will show convergent validity across multiple measures

If the disposition effect is a stable behaviour of an individual, it should be possible to measure it in multiple ways and still find consistent results. The two-index game DE scores measure the disposition effect directly from trading decisions made within its simplified trading environment. The Saxo trading data measures the disposition effect directly from investor's real-world trading in financial markets. Finally, the disposition effect scale completed by Milan and London participants measures the disposition effect using self-reported responses about typical attitudes and decisions during an investor's real-world trading, and relies on respondents' memories. So, these three methods cover a range of measurement types and data sources.

Despite these differences, all correlations between them are significant. The correlation between the DE scores from the game and the Saxo real-world trading is significant, with a medium size ($r = .271$). The correlations between the scale and DE scores from the game are also significant, with medium effect sizes ($.231 < r < .369$). These results support convergent validity of these instruments: they are capturing the same variation in behaviour.

These results support the conclusion that the disposition effect is a stable behaviour which differs between investors, and is driving variation in all three measurements. Since the three types of measurement are all designed to measure the disposition effect, it is reasonable to conclude that the shared variation they are capturing is attributable to variation in the disposition effect between participants. The reason the game, scale, and real-world trading correlate is because people do differ in their disposition effect bias, and this is being successfully measured in these studies.

8.2 DO CUTTING GAINS AND HOLDING LOSSES HAVE TRAIT-LIKE CHARACTERISTICS?

This research question asks whether cutting gains and holding losses also have trait-like characteristics. Therefore, it asks similar questions as the first research question: whether these biases are reliable and display construct validity. These questions are important for the same reasons: intra-individual stability shows that the bias persist over time, and convergent validity shows that the biases are expressed across different situations and methods of measuring it. Intra-individual stability and convergent validity are assessed in the same way as for the disposition effect above.

There is also the additional question of whether the biases are independent of one another. This is key evidence for the question of whether the disposition effect is a monolithic bias or is better described as two distinct trading biases. This is tested by investigating the discriminant validity of cutting gains and holding losses, using the two-index game, the scale, and both simultaneously. Whereas convergent validity implies that the same bias should correlate when measured in different ways, discriminant validity implies that independent biases should not correlate with each other. This would strongly support the claim that they are both measuring stable and distinct biases.

Convergent validity is tested in a similar way as the disposition effect, by comparing scores on the two-index game with scores from the scale (after two factors representing the two biases were extracted from the scale). Discriminant validity is tested in 3

different ways. Firstly, by comparing scores for the two biases within the two-index game. Secondly, by comparing scores for the two biases within the scale. Finally, by comparing scores for the two biases across the game and scale.

Table 8.2 details results for hypotheses relating to the second research question, and then each hypothesis is discussed.

Table 8.2 Findings for trait-like characteristics of cutting gains and holding losses

	Hypothesis	Test	Supports hypothesis	Chapter
2.1	Cutting gains will show intra-individual stability	Correlations between repeated plays of the two-index game	Yes	4
		Variance in cutting gains (PGR) scores accounted for mainly by differences between participants	Yes	4
		Test-retest stability (in the Milan study)	Yes	4
2.2	Holding losses will show intra-individual stability	Correlations between repeated plays of the two-index game	Yes	4
		Variance in holding losses (PLR) scores accounted for mainly by differences between participants	Yes	4
		Test-retest stability (in the Milan study)	Yes	4
2.3	There will be discriminant validity between cutting gains and holding losses in the two-index game	Comparing correlations within and between PGR and PLR over repeated plays.	Partially	4
		Correlations between PGR and PLR over repeated plays after controlling for number of trades	Yes	4
2.4	There will be discriminant validity between cutting gains and holding losses in the scale	Principal components analysis to split the scale into factors representing cutting gains and holding losses	Yes	5
		Correlation between the two extracted factors – testing that the correlation is not significant	Yes	5
2.5	There will be convergent	Correlations between PGR scores and gain factor scores	No	5

	validity for cutting gains between the two-index game and the scale	Correlations between PGR scores and gain factor scale scores	No	5
		Correlations between PGR scores and all-gains scale scores.	Partially	5
		Correlations between PGR scores and selling-gains scale scores.	Yes	5
2.6	There will be convergent validity for holding losses between the two-index game and the scale	Correlations between PLR scores and loss factor scores	Yes	5
		Correlations between PLR scores and loss factor scale scores	Yes	5
2.7	There will be discriminant validity between cutting gains and holding losses, between the two-index game and the scale	Correlations between PGR and loss factor scores, and PLR and gain factor scores. Testing that the correlations are not significant.	Yes	5
		Correlations between PGR and loss factor scale scores, and PLR and gain factor scale scores. Testing that the correlations are not significant.	Yes	5

8.2.1 Cutting gains will show intra-individual stability

The same tests carried out to establish intra-individual stability of the disposition effect were also carried out for cutting gains. Correlations between PGR scores on repeated plays of the game were strong on all 3 studies, and were all significant. All correlations exceeded $r = .5$, except with play 1 on the London study. Typical correlation strength on the Milan and OU studies was about $r = .7$.

Variance in PGR scores attributable to between-participants differences was between 55% and 80%, across the 3 studies. This provides strong evidence that a stable bias which differs between participants drives these scores.

Finally, test-retest correlations in the Milan study were all significant with large effect sizes ($.605 < r < .745$).

8.2.2 Holding losses will show intra-individual stability

The same tests were carried out to establish intra-individual stability of holding losses. The results were even stronger than for the disposition effect and PGR. PLR scores between repeated plays of the game within studies were all significant and very strong, with typical correlation coefficients of $r = .8$. Variance in scores attributable to between-participant differences was similarly strong (72-80%). Finally, test-retest correlations were all significant ($.585 < r < .657$).

8.2.3 There will be discriminant validity between cutting gains and holding losses in the two-index

Discriminant validity concerns whether two instruments which should measure different biases do indeed measure different biases. Biases which are not theoretically related to each other should not be related to each other when they are measured, either with the same method or with different methods. Therefore, PGR and PLR should not be related to one another. Showing this provides evidence that cutting gains and holding losses are two distinct biases which combine to produce an apparent disposition effect.

Discriminant validity was assessed by analysing the pattern of within-bias and between-bias correlations in the two-index game. Between-bias correlations were expected to be low, and certainly lower than the within-bias correlations. Discriminant validity is not tested with a single statistic, but is assessed by the overall pattern which emerges from many correlations.

In the London and Milan studies, within-bias correlations were clearly stronger than between-bias ones overall, giving strong evidence for discriminant validity. The OU study produced different results to the other two studies as all correlations were highly significant, regardless of type. There is still some evidence for discriminant validity between PGR and PLR, since within-bias correlations had larger effect sizes than between-bias correlations on average.

However, the number of trades made is a potentially confounding factor and can lead to PGR and PLR being correlated. To remove this variation, PGR and PLR scores were regressed on the number of total trades made in each game, and the correlation pattern were assessed again when correlating the residuals of these regressions, rather than the raw figures.

After controlling the frequency of trading, between-bias correlations were moderate to weak ($<.300$) and usually non-significant, while within-bias correlations were still strong and all highly significant. In the OU study in particular, the differences between within-bias and between-bias correlations were much clearer after controlling for frequency of trading. So, this provides good evidence that stable and distinct biases relating to cutting gains and holding losses drive these scores, once the frequency of trading has been controlled for.

8.2.4 There will be discriminant validity between cutting gains and holding losses in the scale

The disposition effect scale, completed by Milan and London participants, has 10 items. Some relate to gains and some relate to losses, while some relate to holding positions and some relate to closing them. The aim was to show that the scale could be used to measure cutting gains and holding losses separately, and provide further evidence that they are both stable and distinct biases. It would also establish that further work in scale development is worthwhile and likely to produce reliable self-report measurements of these biases.

Initial analysis of inter-item correlations suggested two main clusters of items based around behaviour towards gains and losses respectively. Principal components analysis formalised this by extracting two distinct factors, one representing each bias. Then principal components analysis using direct oblimin rotation tested the correlation between these factors by allowing them to freely correlate. The correlation was very low ($r = .0073$) and not significant. In both statistical and practical terms, this can be treated as a zero correlation, meaning the two factors are independent of one another.

So, the scale analysis works strongly suggests that even a limited scale based on the disposition effect can be split into two independent factors for cutting gains and holding losses.

8.2.5 There will be convergent validity for cutting gains between the two-index game and the scale

If cutting gains is a stable bias of an individual, it should be possible to measure it in multiple ways and still find consistent results. The two-index game PGR scores measure the disposition effect directly from trading decisions made within its simplified trading environment. The gain factor scores are the result of principal components analysis carried out on the disposition effect scale, completed by Milan and London participants, which measures trading behaviours and attitudes with self-reported responses (from memory). Gain factor scale scores are the result of adding up the scores from the scale items which primarily load onto the gain factor.

In contrast to the hypothesis, there were no significant correlations between PGR scores and gain factor scores, nor between PGR scores and gain factor scale scores. This failure may be due to the difficulty of carrying out exploratory principal components analysis with a small number of initial items. As discussed in chapter 5.5.3.1, the loading of item 3 to the cutting gains factor is not intuitive, and may be the result of a limited number of initial items to use for the factor analysis.

However, constructing scale scores *a priori* on a theoretical basis, rather than those extracted using principal components analysis, provides results that suggest further scale development could result in a gain factor that shows convergent validity.

All four items on the scale which relate to gains were summed to produce 'all-gains scale scores'. These scores have marginal or significant correlations with PGR scores from the game. In addition, the three items which relate specifically to selling gains (as opposed to holding them) were also summed to produce 'selling-gains scale scores'. These scores have significant correlations with all 3 sets of PGR scores. So, there is evidence of a

relationship between measuring cutting gains in the two-index game with some of the scale items.

Since the factor analysis did not produce a factor grouping these items, this does not demonstrate convergent validity for a cutting gains bias, because the cutting gains items cannot be claimed to represent cutting gains when they do not correlate with each other. However, it does provide a solid rationale for studying the cutting gains further, for example why the 'selling-gains' items did not form a factor, and developing the to measure the factor structure of cutting gains with more precision.

8.2.6 There will be convergent validity for holding losses between the two-index game and the scale

The loss factor extracted by principal components analysis was tested in the same way as the gain factor above but using PLR scores rather than PGR scores. In contrast to cutting gains, correlations provided strong support of convergent validity between the scale and game scores. Correlations of the factor scores with PLR on all 3 plays were significant with medium sized correlations ($-.332 < r < -.242$).²⁰ Correlations using the scale scores were similar but not as strong, with one marginal and two significant correlations, with slightly lower effect sizes. However, overall this is a strong result which supports holding losses as a stable bias that drives scores in both the game and the scale.

8.2.7 There will be discriminant validity between cutting gains and holding losses across the two-index game and the scale

Cutting gains and holding losses have already been demonstrated to be distinct within the two-index game, and within the scale. The final set of tests in this section sought to establish that cutting gains and holding losses were distinct biases across measurement

²⁰ The correlations are negative because as the factor scores for holding losses increase, the proportion of losses held should decrease, since losses will be held more and sold less frequently.

methods. So, PGR from the game should not correlate with the loss factor from the scale, and PLR from the game should not correlate with the gain factor from the scale.

The results strongly supported this. Correlations between PGR and the loss factor scores and scale scores were all non-significant with effect sizes around zero. Correlations between PLR and the gain factor scores and scale scores were also all non-significant and around zero.

8.3 DOES COGNITIVE REAPPRAISAL AFFECT THE DISPOSITION EFFECT AND ITS CONSTITUENT BIASES, WHEN TESTED IN RETAIL INVESTORS UNDER CONDITIONS OF GREATER EXTERNAL VALIDITY?

This research question, which relates specifically to the Milan study, set out to test the disposition effect and its constituent biases experimentally, with greater external validity than has been achieved before. An increase in ecological validity was provided by using the two-index game as a trading environment, rather than a simpler trading approach such as Lee *et al.* (1998). In addition, the Milan study recruited experienced traders as opposed to novice participants with no experience of trading, increasing the external validity of the study in a second way.

With this increase in external validity, the research question examined several questions. Do experienced traders trade with a disposition effect in the lab when measured with greater ecological validity? Does cognitive reappraisal reduce the disposition effect under these conditions? Finally, does reappraisal achieve its effect by reducing the tendency to hold losses as opposed to cut gains?

Table 8.3 details tests which were carried out to test these hypotheses, and again the sections following it discuss the results in more detail.

Table 8.3 Findings for the effect of cognitive reappraisal on the disposition effect and its constituent biases, tested in retail investors.

	Hypothesis	Evidence	Chapter	Supports hypothesis
3.1	Investors will show a disposition effect in the two-index game	T-tests on level of disposition effect on each play	6	Yes
3.2	Cognitive reappraisal will reduce the disposition effect	T-tests comparing change in disposition effect between groups	6	Yes
		Marginal modelling of 3 repeated plays, with effect of reappraisal as an interaction	6	Yes
3.3	Cognitive reappraisal will reduce holding losses but not affect cutting gains	T-tests comparing change in cutting gains between groups (testing null effect)	6	Yes
		T-tests comparing change in holding losses between groups	6	Yes

8.3.1 Investors will show a disposition effect in the two-index game

The disposition effect was shown to be present when tested across the entire dataset. A one-sample t-test, testing whether the overall disposition effect differed from zero (zero indicating no bias), was highly significant. Participants sold gains about 75% more frequently than losses. The disposition effect was also demonstrated on plays 3 and 4 analysed separately, with their respective t-tests also highly significant and participants about twice as likely (i.e. a 100% increase) to sell gains as losses. The effect was not significant on play 2 alone, although participants were still about 30% more likely to sell gains than losses.

8.3.2 Cognitive reappraisal will reduce the disposition effect

The experimental design used repeated measures and compared an experimental group with a control group. As expected, cognitive reappraisal reduced the disposition effect relative to the control group. The change in disposition effect between plays 3 and 4 was significantly different between groups, with the reappraisal group showing a relative reduction of 55%.

There were however, potential issues in terms of the initial baseline levels of the groups. To address this, data for play 2 was also included so that three repeated measurements were analysed. Marginal modelling was used to test the effect of play order and group membership across all three plays. The effect of reappraisal was represented in the interaction between play order and group, and specifically in the model's parameter for the reappraisal group on play 4. The interaction overall was significant and the model parameter for the reappraisal group on play 4 was marginal; however, if this test is treated as one-tailed (since it was specifically predicted) then this parameter is also significant.

So, the effect of reappraisal has been tested in two different ways, and found to be significant on each occasion. This is good evidence that reappraisal is still effective, when improvements are made to the external validity of the test from both increasing ecological validity and using representative participants.

8.3.3 Cognitive reappraisal will reduce holding losses but not affect cutting gains

The effect of reappraisal on cutting gains and holding losses were tested in the same way, using repeated measures and comparing the change in an experimental group with that in a control group. As expected, reappraisal produced a significant reduction in holding losses, i.e. an increased willingness to sell losses. At the same time, it was shown that changes in cutting gains were not significantly different between groups.

The combined effect of these two tests is that the effect of reappraisal on the disposition effect can be attributed to a decrease in holding losses, rather than a decrease in cutting gains, and the hypothesis is supported.

8.4 DOES COGNITIVE REAPPRAISAL AFFECT THE DISPOSITION EFFECT AND ITS CONSTITUENT BIASES, BY CHANGING EMOTIONS DURING TRADING, WHEN TESTED IN NOVICES UNDER CONDITIONS OF GREATER EXTERNAL VALIDITY?

This final research question was tested by the OU study. It attempted to build on the findings of the 3rd research question and the Milan study, and retained the emphasis on testing reappraisal with greater ecological validity by using the two-index game to measure trading biases. Participants were adults across a range of ages and backgrounds, tested on the Open University campus. This relaxation of the requirements for participants allowed a larger dataset to be collected, and the greater experimenter control of a study completed in person allowed a wider range of questions to be asked. Although not using retail investors sacrificed some external validity, it was still considered an improvement compared with the narrow cross-section of adults that using student samples provide.

This research question initially covered the same questions as the third question: the existence of the disposition effect, and the effect of reappraisal on the disposition effect, cutting gains and holding losses. However, in addition, the mechanism of reappraisal and its link to emotions experienced while trading was explored. Does reappraisal bring about changes in trading biases and experienced emotions simultaneously? If so, can changes in emotions mediate the changes observed in trading biases?

Table 8.4 Findings for the effect of cognitive reappraisal on the disposition effect and its constituent biases, and the role of emotions in reappraisal's effect, tested in novices

	Hypothesis	Evidence	Chapter	Supports hypothesis
4.1	Novices will show a disposition effect	Confidence intervals for each play, testing whether disposition effect is different from zero	7	Yes, but lower levels of DE than expected, and no overall bias in "professional reappraisers"
4.2	Cognitive reappraisal will reduce the disposition effect	T-tests comparing change in disposition effect between groups	7	No
4.3	Cognitive reappraisal will reduce holding losses but not affect cutting gains	T-tests comparing change in cutting gains between groups (testing null effect)	7	Yes
		T-tests comparing change in holding losses between groups	7	No
4.4	Cognitive reappraisal will reduce negative emotions experienced during trading	T-tests comparing change in each emotion between groups	7	No
4.5	Changes in emotions during trading will mediate the effect of reappraisal	Not tested because there was no effect of reappraisal to mediate	7	n/a

8.4.1 4.1-4.3 The effect of reappraisal on the disposition effect, cutting gains and holding losses, in novices

As with the Milan study, the existence of the disposition effect during the study was established. However, the levels of disposition effect were lower than the Milan study, suggesting that novices were not able to express themselves fully in the more realistic trading setup. In addition, a subset of participants labelled "professional reappraisers" (who chose to carry out reappraisal as some type of financial professional), did not display a disposition effect at the group level.

Contrary to expectations, cognitive reappraisal did not have a significant effect on the disposition effect overall. Those who rated reappraisal as less difficult to carry out did have a larger decrease in disposition effect compared with the control group. However,

splitting the reappraisal group into subgroups based on participants' difficulty ratings reduced sample sizes and statistical power greatly. Consequently, this finding was not significant, so is only tentative.

Since there was no effect of reappraisal on the disposition effect, it was not possible to explain changes in disposition effect in terms of changes in holding losses. There was no effect of reappraisal on holding losses, and as expected there was no effect on cutting gains either. The implications of these results are discussed in the following chapter; however, for now, a summary is that there was no effect of reappraisal on any of the trading biases when testing novices.

There was, however, an effect of reappraisal on perceived responsibility (reported in the manipulation checks section 7.3.1). However, the effect was to increase perceived responsibility when carrying out reappraisal, not decrease it. In addition, participants who rated reappraisal as easier to carry out tended to have larger increases in perceived responsibility. Though the conclusions that can be drawn from these findings are limited by reduced sample size, it would suggest that when participants were able to carry out reappraisal as instructed they felt more responsibility for their decisions, not less as intended. This was the opposite of what was intended.

8.4.2 4.4-4.5 The effect of reappraisal on emotions, and mediation of its effect on the disposition effect

Reappraisal did not have any significant effects on any emotions measured. This included the PANAS positive and negative affect scales, four specific affect scales from the PANAS-X, and five other emotions related to those tested by Summers and Duxbury (2012). Since there was no effect of reappraisal on the disposition effect, nor on any changes in emotions reported, the mediation of the former by the latter was not tested.

8.5 SUMMARY OF ALL FOUR RESEARCH QUESTIONS

The first research question examined trait-like characteristics of the disposition effect by examining its properties. It found that the disposition effect is reliable when measured

experimentally using the two-index game, in the short term and likely the medium term too. It displays convergent validity when compared between the scale, the game, and trading records from financial markets. This demonstrates that the disposition effect does have trait-like characteristics, and can drive variation in individuals' trading behaviour both in the lab and in the field.

The second research question addressed whether cutting gains and holding losses could also be considered as biases, with the additional issue of whether they are independent of each other. There was strong evidence for the intra-individual stability of both holding gains and cutting losses in the two-index game. There was weak support for convergent validity of holding gains, though this may be improved with more extensive scale development. In contrast, there was strong support for convergent validity in holding losses. Finally, there was strong evidence for discriminant validity between cutting gains and holding losses, tested within the game, within the scale, and between the game and the scale.

The third research question asked whether retail investors displayed a disposition effect in the lab, whether it was reduced by reappraisal, and if so whether this effect was driven by a decreased tendency to hold losses. Not only did the Milan study test these hypotheses, but it did so with increased external validity from both the method used to measure trading biases and the participants used. All three hypotheses in this area were supported. There was an initial disposition effect in participants, it was reduced by reappraisal and this was because of the effect on holding losses. So, this research question provides good evidence that reappraisal is still effective when external validity of experimental setups is improved, and that changing behaviour towards losses is the mechanism for reappraisal's effect.

The fourth and final research question tested similar issues to the third question with a similar design, but using a mixed adult sample of novices rather than retail investors. In addition, it sought to test ideas about how reappraisal might affect emotions experienced

during trading, and if these could shed light on how reappraisal affects the disposition effect.

In contrast to the previous study, few effects were found. There was still an overall disposition effect, albeit much lower than previously. Using novices as participants appeared to negate the effects of reappraisal on trading biases. Likewise, there were no effects found of reappraisal on emotions during trading, with only an unexpected increase in perceived responsibility. Since there were no effects of reappraisal on trading biases, nor on emotions during trading, the mediation of the former by the latter was not tested.

9 CONTRIBUTIONS AND DISCUSSION

This thesis has examined four research questions, with a series of hypotheses relating to each one. In fact, this thesis can be considered as two complementary and contrasting themes: whether the disposition effect, cutting gains and holding losses show trait-like characteristics, and whether it is possible to change these trading behaviours in novices and in experts, while increasing ecological validity in experimental testing.

The first two research questions sought to establish whether trading behaviours could be considered stable biases. By examining the intra-individual stability and construct validity of these trading behaviours, it was established that they had trait-like characteristics. In addition, these tests used experimental instruments with improved ecological validity, giving further credence to the view that the trading behaviour measured in the lab is representative of trading biases observed in real world trading.

The latter two research questions attempt to change the expression of those same biases using emotion regulation, but increasing the external validity of the experimental designs used to test this. External validity was increased by improving the ecological validity of the instrument used to measure trading behaviour, and using retail investors as participants in some studies.

The previous chapter summarised the empirical evidence generated in each of the four research questions. This chapter begins by stating the contribution made by this thesis for each question. It continues by exploring the implications of these contributions, and possibilities for future research. It discusses the limitations of the work, and makes suggestions for further future research in response to these limitations. Implications for practitioners are suggested. Finally, the overall conclusions of the thesis are laid out.

9.1 CONTRIBUTIONS

9.1.1 Does the disposition effect have trait-like characteristics?

There are many experimental studies of the disposition effect, and many measures of the disposition effect used. In all these studies, the disposition effect is operationalised by demonstrating a difference in the likelihood of gains versus losses being sold. While this mathematically satisfies the definition of a disposition effect, there is little research on the implied claims that these studies make: that the measures in lab studies are reliable, and that they valid (i.e. that they relate to disposition effects in the real world), and extrapolating from this, that the disposition effect is a stable bias.

If lab studies do not produce reliable measures, then at best they are measuring the disposition effect as state behaviour, rather than trait-like behaviour. If measures are not even stable within a session, it is arguable whether they measure anything at all. So, establishing reliable measures in the lab is a desirable goal.

In addition, many studies also only demonstrate a difference between gains and losses on a group level, for example Lee *et al.* (2008), which this thesis specifically sets out improve on. These studies are not able to claim they are directly measuring a disposition effect for any individual at all, since an individual's trading behaviour is only measured towards gains or losses, but not both. The operationalisation of the disposition effect is assumed to be reliable because the gain and loss groups are drawn from the same population, and it is assumed the gain would have the same bias towards losses as the losses group were they to be tested on losses instead (and vice versa). So, the difference in behaviour between groups, and thus the operationalisation of the disposition effect, is assumed to be reliable. However, such studies do not assess this reliability. To do so, the studies would need to be repeated many times.

Experimental studies are often justified explicitly by reference to the impact of the disposition effect on real-world trading. This follows in the footsteps of the seminal paper by Shefrin and Statman (1985), which attempted to produce a theoretical explanation

for the phenomenon observed in financial markets. Even if not explicitly justified like this, experimental studies on the disposition effect still imply the research is interested in disposition effects that occur in real-world decision making. To phrase this another way, few research papers claim to be interested only in the disposition effect measured in a specific lab setup used in the study. Simplification is usually a practical step to experimentally isolate variables of interest.

This thesis argues that it is justified to treat the disposition effect as a stable trait-like bias that determines a person's trading behaviour in the short term, the medium term, and likely the long term. It can be reliably measured in the lab, with a method that achieves good ecological validity, and it is still reliable when using actual retail investors rather than novices as participants.

This thesis argues that it is a stable trading behaviour: a persistent feature of a person's trading "personality". It is not stable only when using one method to measure, but a participant or investor scoring highly on one measure of the disposition effect should tend to score highly on another. So, the contribution is in demonstrating that the disposition effect has trait-like characteristics, varies reliably between people, and affects different measures of the same trading behaviour. This is an important contribution to the field because previous experimental studies have not established that the disposition effect they measure represents a stable trading behaviour.

Intra-individual stability in the lab has been demonstrated using the two-index game. Although this is a simplification of real world trading, it does incorporate many more realistic features of trading, and give much greater flexibility to the participant in how they make their decisions. Demonstrating intra-individual stability here is a novel finding alone. A disposition effect doesn't need to be measured with the constricted and forced-choice decisions usually used in the lab; it can still be reliably measured despite the random noise that the more realistic features introduce.

Not only has intra-individual stability been demonstrated, but it has been demonstrated very robustly. Three or four repeated measurements were taken from each participant. This was demonstrated convincingly over three independent studies, two studies using retail investors and one study using novices. There can be little doubt that the disposition effect was reliably measured in these studies.

The two studies which used retail investors as participants also demonstrated convergent validity between a self-report scale, the trading game, and real-world trading: three qualitatively distinct types of measure. So, the two-index game combined with the scale and real-world trading records provide strong evidence that the disposition effect has trait-like characteristics, and that it reliably varies between people. The fact that retail investors were used as participants, and that retail investors were also measured trading on financial markets, makes this particularly valuable: it makes a link between the disposition effect used in the lab and the actual trading behaviour of investors.

Overall there is strong evidence that the disposition effect is a persistent feature of decision-making during trading which can be treated as having trait-like characteristics. It can be measured consistently in both novices and experts, with lab measurements, self-report scales, and trading in financial markets.

9.1.2 Do cutting gains and holding losses have trait-like characteristics?

The disposition effect can be operationalised by demonstrating a difference in the likelihood of gains and losses being sold. These “two sides” of the disposition effect, cutting gains and holding losses, can be measured separately in this study for each participant, and the disposition effect is demonstrated by contrasting them.

However, this thesis argues that these two sides of the disposition effect are also stable trading biases in their own right, which determine a person’s trading behaviour towards gains and losses independently. They can also be reliably measured in the lab; improved

ecological validity in the instruments used does not vitiate this stability, and they are stable when using actual retail investors rather than novices as participants.

This second contribution is similar to the first contribution in its motivation to establish the intra-individual stability and validity of measuring trading biases in the lab. As with the first contribution, the claim here is that the biases are not only reliably measured in the lab, but that they are persistent features of a person's trading "personality". We are justified in treating them as biases existing independently of the measure used in any one study. This is an important contribution to the field because previous experimental studies have not established that these biases have been measured with intra-individual stability or validity.

However, in addition to these points, the second contribution also includes convincing evidence that these two constituent biases of the disposition effect are largely independent of each other. This is important because it links back to the first contribution, and explains why the disposition effect is a stable bias. It is argued here that its true nature is actually two stable biases, which correspond to each 'side' of the disposition effect. Though the disposition effect appears to be stable bias itself, which displays convergent validity, the claim here is that this is only the case because its two constituent biases are stable biases themselves which display convergent validity.

To make this contribution, this thesis has demonstrated the intra-individual stability of cutting gains and holding losses in the same way it is demonstrated for the disposition effect. Repeated measurements of the biases were made using the two-index game. Convincing demonstrations of the intra-individual stability of the biases were made in three independent studies: two studies using retail investors and one study using novices.

Convergent validity between the game and scale was also demonstrated in a similar way (after the scale was split into two factors by factor analysis). Convergent validity was strongly demonstrated for holding losses. This was not demonstrated for cutting gains

using the scale factor extracted; however, there was good evidence for a relationship between the trading game and some of the cutting gains items selected on theoretical grounds (rather than statistical grounds).

The additional analysis for this contribution, to show the biases are independent, involved testing discriminant validity. This is the idea that if measures are independent they should not be correlated. This was tested in three ways.

First, using factor analysis the scale was shown to split into two independent factors representing these two biases. Second, using scores from the game, there was good evidence across the three studies that scores for cutting gains (measured by PGR) were correlated weakly with scores for holding losses (measured by PLR). This evidence was particularly strong after controlling for frequency of trading.

Finally, data from the game and scale were used together, as they were for establishing convergent validity. However, opposing biases were compared instead of matched biases, so cutting gains from the game was compared with holding losses from the scale (and vice versa). Using all three methods, the biases were found to be independent of one another.

Overall there is strong evidence that the two sides of the disposition effect are effectively separate trading biases, which can be reliably measured in the lab. In most experimental work, they have been combined into the disposition effect, since this allows a relative score for their 'disposition effect bias' to be produced. However, this thesis clearly shows that this analysis is incomplete.

9.1.3 Does cognitive reappraisal affect the disposition effect and its constituent biases, when tested in retail investors under conditions of greater external validity?

The first two contributions provided evidence for the trait-like characteristics of the disposition effect, cutting gains and holding losses. Having established these, the third

contribution is concerns changing the expression of those trading biases using emotion regulation, while increasing the external validity of the studies used to test this.

The two-index game significantly improves on previous lab methods to measure the disposition effect, while the use of retail investors demonstrates these effects in retail investors, rather than the novices usually featured in experimental work. The Milan study employed both improvements to verify that the disposition effect occurs, and that cognitive reappraisal can be used to reduce it in retail investors.

This builds directly on Lee *et al.* (1998) who used a form of reappraisal to manipulate the disposition effect. It also echoes others who have demonstrated the ability of reappraisal to change other related biases such as loss aversion (e.g. Sokol-Hessner *et al.*, 2009), and it re-affirms the broader literature on reappraisal as an effective tool to improve decision-making. However, what previous studies lacked is a demonstration that reappraisal can be applied beyond artificial lab environments, and that it can be effective with experienced participants rather than novices.

So, the main contribution here is to demonstrate reappraisal's effectiveness as a de-biasing tool, using an experimental setup much closer to real world conditions, with retail investors as participants. This reinforces the findings of previous studies: as discussed in the first contribution above, most researchers are ultimately motivated to research the disposition effect in the lab because of its effect in the real world, starting from the seminal work of Shefrin and Statman (1985).

This motivation also applies to the use of reappraisal. It is interesting to note that reappraisal is effective in the lab, but better to demonstrate that it will be effective in the real world, and the improvements in external validity provide stronger evidence that this is the case. With the current concerns about replicability in psychological science, more work showing that psychological manipulations in the lab have external validity is welcome.

The Milan study also produced a novel finding, beyond its improvements in external validity. The experimental design allowed cutting gains and holding losses to be measured separately, and the within-participant design allowed changes in each to be monitored. This led to the finding that the decrease in the disposition effect, which cognitive reappraisal produces, can be attributed to a decreased tendency to hold losses (i.e. an increased willingness to sell losses) rather than a decreased tendency to cut gains.

This ties in well with the second contribution which showed that the disposition effect can be de-composed into its two constituent biases. In effect, previous studies have not told the whole story when focussing on the disposition effect as a difference between behaviour towards gains and losses. This thesis fills in the gaps by showing that holding losses is a stable bias in its own right, and that it is behaviour towards losses that drives the effect of reappraisal on the disposition effect.

9.1.4 Does cognitive reappraisal affect the disposition effect and its constituent biases, by changing emotions during trading, when tested in novices under conditions of greater external validity?

The final contribution of this thesis is to raise interesting questions about the difference between studying novices and experts, and how this difference was affected by increasing external validity in the studies used. The Milan and OU studies were conceptually identical, with very similar procedures which both attempted to increase the external validity of previous results. However, they differed in the type of participant used and this appears to have produced the contrasting results.

The Milan study replicated and improved upon previous findings, showing that reappraisal reduced the disposition effect by reducing holding losses. In contrast, the OU study demonstrated that if ecological validity is increased when measuring the disposition effect, but novices are used as participants, cognitive reappraisal is no longer effective in reducing the disposition effect. Nor did the OU study find an effect of

reappraisal on the other variables tested, including cutting gains, holding losses, and various measures of emotions experienced during trading.

Given the positive results of the Milan study (and many previous studies) where reappraisal was effective, this lack of effect is best explained as the result of the failure of reappraisal to have any effect on novices in this setup, rather than the lack of effect of reappraisal in general. The same applied to the involvement of emotions during trading as a mechanism for reappraisal's effect, which the OU study sought to test. If reappraisal had affected the trading biases but not emotions during trading, this would be evidence that emotions are not involved in mediating reappraisal's effect. However, given the null effect of reappraisal on trading biases too, all that can be said is that the expected effect of reappraisal did not occur.

This contribution raises two main issues. The first relates to using novices when attempting to study psychological biases with greater ecological validity. While improving ecological validity is desirable, novices' lack of expertise may interfere with their ability to carry out tasks in ecologically valid studies, even without the effect of biases. Thus, testing the experimental manipulation of biases this way is problematic.

The second issue concerns generalising from lab studies using novices, to real world settings which require expertise or experience. These results suggest that de-biasing techniques such as cognitive reappraisal may be less effective when implemented in real world settings where expertise or experience are a factor, than they are when implemented in the lab using novices. This calls into question the implications of previous findings about the effect of cognitive reappraisal on trading biases. It suggests caution should be used when extrapolating from positive results in the lab with novices, to the same phenomenon in real world behaviour.

9.2 IMPLICATIONS AND FURTHER RESEARCH

9.2.1 Trait-like characteristics of the disposition effect, cutting gains and holding losses

Since the first two contributions are theoretically similar, and share many implications, they are discussed together here. The disposition effect has been shown to have trait-like characteristics, in its stability over time and convergent validity across different measures. However, it is also argued that the disposition effect can be decomposed into (at least) two constituent biases. The thesis provides strong evidence that cutting gains and holding losses have trait-like characteristics too, which can reliably be measured in the lab and which drive trading behaviour in the real world. The disposition effect is a stable bias, but this is simply the combined result of the first two biases. If cutting gains and holding losses are both reliable and both show convergent validity, then so will the disposition effect.

If cutting gains also splits into constituent biases as speculated, then an analogy can be drawn with the constituent biases of the disposition effect. Although the two-index game does not produce any measurements for constituent biases of cutting gains, if those constituent biases are reliably elicited by the game, then the measurement of the combination of those behaviours towards gains (measured as PGR) will also be reliable. PGR and PLR are measured separately in this thesis, while the proposed constituent biases of cutting gains are not. Likewise, in many studies where only the disposition effect is measured, PGR and PLR are only measured in combination. However, whether underlying biases are measured separately or only in combination, their underlying psychometric properties will remain the same.

The implication of these conclusions about the disposition effect supports existing work on the disposition effect, and previous results in the lab. As noted, many experimental studies measure 'the disposition effect' operationalised in some way, but implicitly (or explicitly) aim to measure a stable bias in trading behaviour. This thesis supports the

assumption made in these studies: it is justified to study the disposition effect in the lab and generalise to real world trading.

This thesis demonstrates that by using a relatively brief and simple trading task, individual disposition effects that relate to real world trading can be reliably measured in the lab. Using this kind of instrument (one with greater ecological validity) would improve future experimental work on the disposition effect and other biases. In effect, it suggests an aspiration for future work to also use more realistic measures of the trading biases being studied.

Although this thesis does give support to other experimental studies of the disposition effect, it may be that its success in establishing convergent validity is limited to lab measures which have high ecological validity, as the two-index game does. A question for further investigation is whether more basic measures of trading decisions also capture useful information about trading biases. For example, do the measures used by Weber and Camerer (1998) also correlate with real-world trading decisions? Do the more simplified ones from Lee *et al.* (2008)? This could be established by comparing more simple measures to the two-index game, to the scale, or to actual trading records.

Many experimental studies of the disposition effect have assessed the disposition effect by comparing groups. However, this thesis has shown that individual disposition effects can be reliably measured too. So, another aspiration for future research is to study participants' individual trading behaviour, rather than losing information by inferring biases between groups.

The intra-individual stability of the trading biases established here applies to the short term: within the same hour for most of the data, and no more than a few weeks for the test-retest data from the Milan study. A wider claim would be that these trading biases are stable behaviour much like a personality trait, over much longer durations. The convergent validity with the scale provides some evidence that stability is longer term, since participants answered this scale in relation to their historical trading patterns.

However, longitudinal studies would provide stronger evidence on the persistence of individual differences in these trading biases.

Other research has investigated the link between demographic characteristics and the disposition effect, such as Dhar and Zhu (2006). Longitudinal research could answer whether it changes as demographic characteristics change, or whether the link with demographic characteristics is driven by other factors, and the biases are stable in the longer term. For example, perhaps older investors (lower disposition effect) tend not to begin trading until a later age, so it appears that disposition effects decline with age. However, it is possible that these people would have lower disposition effects if measured when they were younger, but they are usually not trading when younger, so are not sampled in field research in financial markets.

In addition to the results about the nature of the disposition effect, this thesis has also made important contributions about cutting gains and holding losses being stable and distinct biases. The implications for researchers here mirror those of the disposition effect. It is important for researchers to use instruments which can measure these biases reliably in the lab and realistically, and this is an aspiration for future research.

In fact, a simple suggestion is simply to ensure these biases are measured at all. The distinction found between the two biases means it is important for researchers to measure them separately whenever possible. It's been discussed above that there is no disposition effect in reality – it is the product of these two separate and independent biases. Therefore measuring only the disposition effect misses information about a participant or investor's trading behaviour: it misses information about how participants with similar disposition effects may have quite different trading behaviour when analysed as two biases. Therefore, future experiments, both on the nature and manipulation of the disposition effect, should aim to measure both sides to understand which is causing a difference between participants.

This distinction also has implications for explanations of the disposition effect. Shefrin and Statman's proposed explanation based on prospect theory was that being in a gain or loss frame changes the subjective utility of further gains or losses. The value function from prospect theory is concave for gains and convex for losses (sometimes called the reflection effect). This means that changes in subjective utility when moving the position back to the reference point are valued more than further losses or gains. In other words, there is a diminishing marginal effect of further gains and losses. This results in investors preferring to sell gains but hold losses, *ceteris paribus*.

The value function in prospect theory transforms a prospect x as follows, depending on whether it is framed as a gain or a loss:

$$v(x) = x^{\alpha} \text{ if } x > 0$$

$$v(x) = -\lambda(-x^{\alpha}) \text{ if } x < 0$$

The difference in steepness between the curves comes from the loss aversion parameter, λ . However, the rest of the equation is determined by the single parameter α in the equation for the value function (which is why it is a reflection – it has the same shape for both gains and losses, but they are the mirror image of each other). Crucially, the difference in subjective utility between returning to the reference point, and a further gain or loss, is the result of α . So, it is α which should drive the disposition effect on this explanation. The lower α is below 1, the greater the deviation in the value function from a straight line, and the more biased the investor should be.

The implication is that cutting gains and holding losses should be correlated within individuals, since both biases are driven by the value of α . The second contribution relating to the independence of cutting gains and holding losses, shows that this is not the case. Therefore, an explanation of the disposition effect based on prospect theory does not appear to be consistent with the results found. This supports previous work questioning an explanation based on prospect theory, such as Kaustia (2010), who found that the probability of selling jumps as at the break-even price.

Future research could investigate the failure of prospect theory to account for the results here, other explanations which could take the place of prospect theory, and how or if they can be reconciled with prospect theory at all. For example, an alternative theory for why people hold losses could simply be that people dislike losses compared with gains, and combined with mental accounting that does not treat a loss as incurred until it is closed, this produces the aversion to close losses. This is consistent with Kaustia's results, which showed that a trading position's profit or loss is treated more like a binary characteristic. People were relatively insensitive to the magnitude of the profit or loss, but responded strongly to whether it is a gain or a loss. So, for example, a loss of £100 is treated similarly to a loss of £500 than a profit of £100, even although in relative terms it is £400 different from the former and only £200 different from the latter.

This kind of explanation would be consistent with the "system 1 / system 2" model of how emotions can affect decisions. People are averse to losses. Combined with falling prey to mental accounting, perhaps they avoid selling losses simply to avoid the pain of psychologically feeling losses. So, investors allow the fact they have a loss *per se* to cloud their judgement, and give this fact more importance than it warrants in objective terms. For the other side of the disposition effect, this logic would be reversed: the pleasure from selling gains should encourage selling them, over and above the objective value of what those gains are worth. This "emotional logic" closely follows the explanation tested in Summers and Duxbury (2012).

The contribution that cutting gains and holding losses are distinct biases also raises the possibility that different psychological mechanisms may underlie them. Cutting gains is about attraction to positive emotions, whereas holding losses is about avoiding negative ones. Put another way, investors who are not biased in either way are able to delay gratification from selling gains, and accept the pain of selling losses.

These seem to be qualitatively different situations, and it is possible that different brain mechanisms underlie each behaviour. In fact, put this way, finding that the two biases

are independent is not actually very surprising – the two biases can occur in quite different situations. Future research can build further evidence that these two biases are distinct, for example investigating possible neural or physiological correlates of each bias, and the independence these correlates.

A further implication of this work is that the disposition effect scale is ripe for further development. The scale was only 10 items long, yet appears to have validity in measuring the disposition effect, without any further selection of items needed. The correlations were significant and medium-sized ($r \sim 0.3$). This kind of effect size demonstrates a relationship between two variables, but the amount of shared variance is still low ($\sim 10\%$). Constructing a scale with a larger sample of items should produce stronger correlations that improve on this by including more detailed items; the results so far suggest such work may be fruitful.

Implications for scale development for the two constituent biases mirror those for the disposition effect. Given the initial success of developing two subscales which measure one constituent bias each, it appears that there is scope to develop this into a longer and more accurate scale. However, for the case for doing so for the constituent biases is even greater. Given the 10 initial items, the selection of items for the subscale was very limited, and this may have led to the failure to establish convergent validity for cutting gains, for example. Further scale development can show whether convergent validity can be improved with a more extensive range of gain items. It could also answer the question of whether cutting gains is a unitary bias, or has its own constituent biases.

9.2.2 The effect of cognitive reappraisal on the disposition effect, its constituent biases, and emotions during trading. Tested in experienced traders and novices under conditions of greater external validity

The second two contributions both deal with the effect of cognitive reappraisal on trading biases. In addition, some of the implications are the result of contrasting the two studies. Given these points, it makes sense to discuss their implications together.

9.2.2.1 Reappraisal as a de-biasing method

Reappraisal was found to reduce the disposition effect in experienced traders using a more realistic method of measuring it. It is plausible that this instruction may work differently in students compared with retail investors. However, an aim of this study was to explore whether cognitive reappraisal could be used as a technique to help retail investors improve their financial decisions, so it was important to test whether people experienced in stock market trading would react to such an instruction in the same way as naïve participants. The results here suggest that this is the case, and reappraisal can have similar effects in both types of participant, as shown in novices in Lee *et al.* (1998), and in the Milan study here.

We can conclude the previously found effect of reappraisal was not caused by the simplicity of the decision-making environment used in that research. The two-index game allows much greater flexibility for participants in how and when they make their trading decisions. It is more cognitively challenging to play since the price changes continuously, there is predictive information about future prices, and this predictive information also changes continuously. Despite this, a disposition effect was still found and reappraisal reduced this.

Reappraisal is also effective with retail investors, so a lack of expertise in novice participants can also be ruled out as an explanation for reappraisal's effectiveness. The previous effect of reappraisal does not appear to be a product of "naïve" decision-making which experts would not engage in.

The results here suggest that it is worthwhile testing whether this form of reappraisal can be effective in real-world trading on financial markets, with experienced participants trading their own capital. The Milan study was a compromise between the previous experimental tests on the disposition effect and a real-world test. Previous studies have been carried out with artificial trading tasks and unrepresentative samples. However, an ideal study would use experimental studies of investors' decisions when trading on

financial markets, and randomly allocate retail investors to an experimental (employing reappraisal) and a control group. This would be very similar to a field study using secondary data, the only difference being the reappraisal instructions given to the experimental group.

Unfortunately, such studies pose significant practical problems. Finding willing participants and convincing them to try a technique when trading their own capital, when that technique is not already fully tested, may be difficult. The design would also require ensuring that investors implementing reappraisal were doing so correctly and consistently. This would also have to be maintained for a sufficient period to produce a trading record large enough to measure trading biases. Most traders do not trade as frequently as in the two-index game, meaning field studies often must draw data from many years for trading biases to be detectable over random noise, allowing a disposition effect to be accurately measured for participants. Asking investors to persist with an experimental approach over long time scales could also be difficult.

Researchers' potential liability for any economic losses sustained might be a practical constraint issue. Participants may worry that reappraisal could actually worsen their trading, and give up if they suspected this. Unfortunately, given a sufficiently large number of investors it is inevitable that some of them would lose money during the study, and this may attribute this to taking part in the study, rather than themselves or simply bad luck.

Of course, these difficulties were an important consideration in using the two-index game in this thesis as a proxy for a trading environment, rather than carrying out a field experiment. However, given the results of the Milan study, there is now more evidence to justify testing reappraisal in financial markets.

9.2.2.2 *The role of emotions in the disposition effect, and cognitive reappraisal's effect*

A novel finding in this thesis is that the effect of reappraisal on the disposition effect can be attributed to a decrease in holding losses, rather than a decrease in cutting gains. Reappraisal had negligible effect on the tendency to close gains. Since the second contribution established that cutting gains and holding losses are distinct biases, we can state this more succinctly: when testing experts, reappraisal reduces holding losses in experts, but does not affect cutting gains.

The Milan study did not investigate the mechanism of reappraisal, but only the effects of its application. However, these results are consistent with the idea that cognitive reappraisal is effective because it changes the emotions that participants experienced during decision making. This is *prima facie* how a form of emotion regulation would affect a decision-making bias, especially a bias which is suspected to be influenced by emotions to begin with.

Since the Milan study found that changes in holding losses were the main driver of the change in the disposition effect, this implies that reappraisal changes or reduces emotions associated with losses, rather gains, and these emotions are presumably negative ones. If experiencing negative emotions results in holding losses more, then any method of reducing or controlling these emotions should result in lower holding of losses and a lower disposition effect overall.

This is consistent with previous research such as Sokol-Hessner *et al.* (2009) who found that reappraisal was effective in reducing loss aversion. Summers and Duxbury (2012) also found the disposition effect was associated with changes in negative emotions (regret), based around the framing of losses. Kaustia (2010) found a qualitative difference between how gains and losses are traded. However, this study did not look at emotions directly, and could not distinguish between whether it was the negative effect of losses, or the positive effect of gains (or both), which was driving this difference.

Unfortunately, the Milan study did not provide any direct evidence that reappraisal operates by reducing negative emotions toward losses, since measures of emotions were not included in that study. These measures were included in the OU study, since that study aimed to provide more direct evidence that cognitive reappraisal reduces the disposition effect by reducing negative emotions associated with trading losses.

Unfortunately, the OU study did not find an effect of reappraisal. However, the results of the Milan study suggest that further work to investigate the mechanism or mediating variables for cognitive reappraisal is worthwhile. Given the results of the OU study, these studies could either use simple trading instruments with novices, or more ecologically valid ones with experts (this is discussed further below).

9.2.2.3 The effect of reappraisal on emotions during trading

The purpose of measuring emotions in the OU study was to investigate if changes in emotions mediate the effect of reappraisal; however, since there was no effect of reappraisal on trading behaviours, there was nothing for emotions to mediate. In addition, there were no effects of reappraisal on emotions, so there were no changes in emotions to do the mediating. Therefore, no firm conclusions can be drawn about whether emotions mediate reappraisal's effect.

The only significant effect of reappraisal was an increase in participants' perceived responsibility for the outcome of their decisions. This was the opposite of what was expected. Reappraisal was intended to allow participants to distance themselves from their decisions, which should decrease the emotions felt during those decisions and reduce trading biases. Instead, when trading someone else's money, participants appear to feel *more* responsible than if they are trading for themselves. Although there were no significant effects of reappraisal on emotions, the direction of the effects was that the control group improved relative to the reappraisal group; this is consistent with reappraisal making participants *more* emotionally involved, not less.

Further research can attempt to establish the effect of reappraisal on emotions during trading. Given the success of the Milan study, it makes sense that reappraisal would alter emotions during trading; however, this has not been established directly in these studies. In a similar way to the design of Summers and Duxbury (2012), one experiment could measure the impact of experimental manipulations on emotions, while another could measure its impact on trading decisions. In combination, an effect of emotions on trading decisions can be inferred.

9.2.2.4 The specific reappraisal instruction used

The results of the OU study also shine a spotlight on the variety of ways in which reappraisal can be implemented. Cognitive reappraisal is a technique people can apply, but the specific way it is implemented can have dramatically different outcomes. Cognitive reappraisal as a technique in general has no predictable effect, but qualitatively different methods of reappraisal can produce quite different effects on emotions. This was demonstrated by the OU study, where an instruction intended to decrease perceived responsibility increased it.

Furthermore, the relationship between cognitive reappraisal and emotions experienced is also complex. While applying cognitive reappraisal was expected to alter the emotions experienced, and help reduce bias in trading decisions, this did not occur there. This may be related to the increase in perceived responsibility rather than a decrease, which was expected to mediate the decrease in emotions experienced.

Another important implication is that the target of reappraisal (i.e. who the participant imagines they are trading for) affects their emotional experience during trading. The reappraisal instruction used in this thesis was based on the one used by Lee *et al.* (2008) where they asked participants to imagine trading as a broker, for clients. This was expected to distance participants from the decisions being made, and in doing so, be less affected by emotions when making decisions. However, when trading “for” a

family member or acquaintance, emotional experience appears to be increased; this was the opposite of the effect intended and may have contributed to the null findings.

Future research could examine how different types of reappraisal affect emotions experienced during trading. For example, Sokol-Hessner *et al.* (2009) used a cognitive reappraisal instruction which was also intended to reduce emotions during decisions. Indeed, they show that a reduction in skin conductance response (SCR) when exposed to losses is correlated with a decreased loss aversion in decision making. However, their instruction focussed more on the attitude toward the decisions being made, rather than relying on a change in perspective (i.e. who participants imagined being, and trading for). This may have been more effective in changing the attitude towards decisions, and therefore changing the emotions experienced during decision making. It would be very interesting to replicate the Milan and OU studies using this kind of reappraisal instruction.

9.2.2.5 The effect of reappraisal in retail investors compared with novices

Combining the results of the Milan and OU studies, reappraisal reduced the disposition effect in the Milan study, but the OU study showed no effects of reappraisal on the disposition effect, nor on emotions during trading. This raises the question of what difference between the studies resulted in reappraisal working in one but not the other.

One clear difference is that in the Milan study, the participants were retail investors, while in the OU study the participants were novices (a range of adults based on a university campus). It is possible that greater ecological validity nullifies the benefits of reappraisal for novices, but not for experts. So, this thesis adds to a topical debate on replicability of psychological research. It suggests that experimental effects on trading biases, found in psychologically simplified and controlled conditions, should be treated with caution until further research can test these effects in more ecologically valid settings.

If researchers are interested in the validity of those effects in real world settings, which is usually the case, then testing in the lab should be only the first step. What appears like a robust finding in the lab may be more difficult to implement in real-life, or simply in more realistic conditions, or may only generalise to specific populations.

The results are particularly interesting when compared with existing literature on the effect of reappraisal on the disposition effect. Reappraisal previously reduced it, when carried out in artificial trading tasks with novices. The Milan study still found a reduction when external validity was improved, both in the measurement instrument and participants used. However, when ecological validity was improved but the participants reverted to novices in the OU study, there was no effect. So, while improving ecological validity is desirable, as discussed above, it may make studies using naïve subjects even less generalizable to the real-world behaviour of experts, which researchers are often ultimately interested in.

This conclusion about the contingent effect of ecological validity also raises the question of why this difference would occur. It is well-known that there are systematic differences between how novices and experts make decisions. Further studies could examine the process of decision-making in experts and novices in trading, and how this might interact with the process of reappraisal they were asked to implement.

Another difference between the studies could be the cognitive load placed on the two sets of participants. This difference could arise from the cognitive demands to play the trading game, or the cognitive load of carrying out this type of cognitive reappraisal, or a combination of both. For retail investors, the two-index game is actually a simplification of their usual trading environment. In contrast, for novices the two-index game was a significantly challenging task in its own right. (This was noted by the researcher speaking to participants during and after participation, and in some of the free-response questions at the end of the questionnaire. Unfortunately, there were no quantitative measures built into the study about difficulty of playing the game per se, as opposed to

the difficulty of reappraisal). So, the two-index game is relatively more difficult for novices to play, and places increased cognitive demands on them before any reappraisal intervention.

At the same time, the cognitive reappraisal instruction used in these studies is a cognitively demanding task itself. It requires participants to continuously imagine they are trading for someone else, or as someone else. Without practising this, cognitive reappraisal itself could demand many cognitive resources throughout the duration of the game. It is possible that limits in cognitive capacity prevented reappraisal being carried out effectively by novices, as they were already struggling simply to play the two-index game effectively.

There is some support for this in the data. The more difficult those participants found reappraisal (measured by their rating of its difficulty), the less effective it was. Participants who rated reappraisal as hard or very hard had an increase in their disposition effect, relative to the control group, while participants rating reappraisal very easy did better than the control group. So, this interpretation has some evidence from the data, but cannot be held with conviction because of the decreased power from splitting the reappraisal group into five subgroups. Replication with increased sample size would help clarify these issues.

Even for those who found reappraisal easy to implement, cognitive load may have put limits on how much they could improve, since they had to split their attention between implementing reappraisal and playing the game. If this is correct, participants who found reappraisal very easy had a lower demand on their cognitive resources, allowing reappraisal to have a more obvious beneficial effect on their trading decisions.

This hypothesis about cognitive load suggests many avenues for future research. Further research could investigate the cognitive load of both the instrument used to measure decisions, and the form of emotion regulation used by participants. If the novices in this study were only impeded because of their unfamiliarity with trading, this could be tested

with further research. Reappraisal should become effective once novices can play the game easily, which can be investigated by habituating novices to the demands of the trading environment. Research using longer duration participant involvement, perhaps using longitudinal designs, could test whether this explanation is valid.

Alternatively, the demands on cognitive load may come mainly from carrying out cognitive reappraisal. Cognitive load was not directly measured, only the difficulty of implementing reappraisal. Future research could investigate the cognitive demands of reappraisal as a primary aim. Perhaps there are other methods of emotion regulation which are cognitively less demanding. Other forms of emotion regulation which do not require constant attention to apply should be more effective in novices.

9.3 LIMITATIONS AND FURTHER RESEARCH

Like the implications section above, the first and second contributions have been grouped together, and so have the third and fourth contributions. Since each pair share much of the same underlying theory, they share many of the same limitations too, and are more parsimoniously discussed together.

9.3.1 Trait-like characteristics of the disposition effect, cutting gains and holding losses

Playing the two-index game involves a specific type of trading. By necessity it measures trading within a brief period, so it can capture trading behaviour within the lab. To generate sufficient trades to measure the trading biases of participants, the two-index game is designed to elicit rapid decision-making and trading in response to a very rapidly changing price index.

However, many investors trade over much longer periods of time. For example, investors may consider selling a stock over days, weeks, or even months. So, there is a question about how similar trading behaviour in the two-index game would be to more deliberative, long term trading. Of course, the same limitation applies to most experimental studies of trading behaviour. Arguably other studies suffer from it more,

given the improvements to ecological validity which this thesis makes (discussed early in the thesis).

Regarding the nature of the disposition effect specifically, this thesis provides good evidence that the disposition effect as measured by the two-index game is shared with investors' long-term behaviour, from the correlation between the game and trading records. In addition, measurements from the game correlate with the self-report scale about trading behaviour in financial markets. However, these findings are not definitive about the extent to which the two-index game captures other types of trading behaviour. In addition, reappraisal was only tested when trading in the game, and not in real-world trading.

What both short-term and long-term trading have in common is that at some point in time, investors must consider selling gains and losses. No matter how long a decision takes, and the amount of time available to make that decision, there is still a point in time where the decision is made. An assumption of all work in this area is that the psychological processes which occur when considering selling stocks are the same regardless of the timeframe, and this is also the view adopted here. However, future research could seek to support this rather than assume it, and build on the conclusions here about the duration over which the disposition effect, cutting gains and holding losses measured in the lab are representative of trading in other environments.

The scale was also limited, since there were few initial items used. Despite this several positive results were found, so this seems like more of an opportunity than a limitation. Further scale development, for example as described by Hinkin (1998) will hopefully produce a scale which is even more closely correlated with the disposition effect seen in trading and build on the initial success demonstrating convergent validity here.

The limitations of the scale were more apparent for the constituent biases rather than the disposition effect itself. It is very encouraging that the two subscales demonstrated independence from one another, and impressive that convergent validity was

demonstrated for holding losses between the scale and the game. However, a longer scale would clearly be desirable.

Compared with the disposition effect, there is more to gain from further scale development for the constituent biases. After removing items which did not load clearly onto one factor or the other, there were only 6 items remaining, and 3 items in each subscale is the very minimum possible to have a scale at all. So, the construct validity of these subscales was very likely hampered by a lack of items to construct a scale from.

Carrying out factor analysis with a larger number of initial items should allow a more robust scale to be produced that correlates more strongly with the game. This would improve the accuracy of the scale as a diagnostic tool; at present the strength of correlations only allows the scale to be indicative of someone who may have a higher than average bias.

The limitations with cutting gains were greater than for holding losses, and suggest this bias will especially benefit from further research. It was not possible to establish convergent validity between the scale and the two-index game for cutting gains. However, the cutting gains component which emerged from principal components analysis was not theoretically intuitive, since it included item 3 which relates to loss aversion rather than cutting gains specifically. The factor loading of this item was lower than the other two items for cutting gains (items 1 and 9), and lower than all 3 of the items loading onto holding losses. Given that item 3 appears to describe loss aversion, it is also curious that this item did not load onto the holding losses factor.

In effect, only two cutting gains items clustered strongly (items 1 and 9). So, it is possible that the failure to find significant correlations is attributable to limitations in the range of items that were included in the original scale. Further research should explore self-report scales relating to gains, by including a larger range of items initially.

An alternative cutting gains scale, using groupings based on theory rather than the result of the factor analysis, created a gains scale which did correlate with PGR game

scores. However, it cannot be used to support convergent validity, since the items chosen did not group together during factor analysis of the initial scale. Why the empirical results of the factor analysis did not match the theoretical grouping is an open question for further investigation. One possibility discussed in chapter 5 is that cutting gains is not a unitary bias but is composed of several constituent behaviours itself. This would result in each constituent of cutting gains correlating with cutting gains as measured by the two-index game, but those constituents would not correlate with each other, and is consistent with what was observed.

These results again strongly suggest that developing a longer scale from a wider range of initial items, to capture these constituent biases, would be a worthwhile next step. This has the potential to explore: how participants interpret item 3 and how this item relates to both cutting gains and holding losses; why the theoretical grouping of cutting gains items did not produce a factor together; and whether the factor structure of cutting gains is better represented as two distinct behaviours.

9.3.2 The effect of cognitive reappraisal on the disposition effect, its constituent biases, and emotions during trading. Tested in experienced traders and novices under conditions of greater external validity

The third and fourth contributions tested the effect of cognitive reappraisal on the trading biases, so are also discussed together.

As discussed above, the two-index game encourages rapid decision-making to produce estimates of a participant's disposition effect. A limitation is that this may limit the ability to generalise from trading in the two-index game, to long-term trading on financial markets. It is possible that, if the process of decision-making is qualitatively different between short-term and long-term horizons, it is possible that reappraisal could change decision making in one context but not another. Therefore, the positive result found in the Milan study may not necessarily apply to long-term trading.

9.3.2.1 Practical problems in the Milan study

An unexpected finding in the Milan study was that the reappraisal group had a higher disposition effect on play 3 compared to the control group. The two groups were randomly assigned, and there was no difference in procedure between them until receiving the cognitive reappraisal instruction immediately before play 4.

One explanation discussed in chapter 5 is mean reversion: within-participant variation in scores on play 3 happened to align with group membership. Since it is unsystematic variation, it did not appear on play 4 which brought the group means close together. An alternative explanation is that the groups differed systematically in their existing disposition effect, and that it is the real effect of reappraisal that resulted in the decrease in score for the reappraisal group on play 4.

To address the concerns about mean reversion, the analysis also included data from play 2 to allow more information about typical disposition effect of each participant. The earlier contributions supported this approach, since it has been shown that disposition effect scores from the two-index game correlate strongly and are likely to be the result of individual differences in trading behaviour. This multilevel model confirmed the original result. Although the reduction in DE attributable to reappraisal was smaller than when comparing groups using a t-test, the main reason for the interaction between play order and group was still the reappraisal group's reduction in DE scores on play 4. So, this provides some support for the explanation based on intrinsic differences in disposition effect rather than mean reversion.

There were some other data issues where mitigation with additional tests was not possible, resulting in sample size being lower than planned and group sizes being unbalanced. Both small sample size and unbalanced groups were ultimately consequences of the study being completed online, and requiring participants to follow the instructions sent to them by email. So, a lack of researcher control over how participants completed the study had a big effect on the quality of data produced.

Although 117 participants were recruited in Milan, just fewer than half this number completed the study online. While participants are of course free to leave at any time during any experiment, in practice this is much less likely when participants take part in person.

Not all participants who completed the study followed the full protocol, for example not completing the surveys in the study. This was especially an issue for the reappraisal group, since the requirements for being included in the reappraisal group were more stringent than the control group: reappraisal participants needed to complete the final survey so that they read the reappraisal instruction. Following the instruction there was a manipulation check to ensure that they had read the instruction and were going to implement the instruction. Participants who did not follow the protocol were moved to the control group, which is why the groups became unbalanced.

One simple extension of this work would be to replicate the Milan study with larger samples or more closely controlled experimental designs, which would hopefully avoid or mitigate these problems. A larger sample size would make up for attrition in participant numbers over the study, leaving good statistical power despite a reduction in initial sample size. In addition, a larger sample should also have limited the impact of random differences in baseline: random differences between groups are likely to become trivial as sample sizes increase and random differences between specific individuals balance out. If possible, a more closely controlled experimental setup would hopefully avoid or mitigate the problems of participants not completing studies and not following the protocol properly.

Of course, the difficulties encountered here highlight a persistent problem in psychological science. To do lab studies in controlled studies is easier than doing field studies. So, we have a situation where studies are usually intended to have external validity, but well-controlled research is valued more highly by scientific journals.

The Milan study sought to improve external validity by recruiting expert participants, and using a more realistic (and subsequently longer) method of measuring trading biases. Both these alterations resulted in practical problems obtaining data, and carrying out a rigorous experimental study. However, a claim of this thesis is that despite the difficulties, this sort of research is still very much worthwhile. Although confidence in its results comes with more reservations, the importance of its results is greater.

9.3.2.2 Contrasting the Milan and OU studies

The implications of the difference in results are discussed above. The interpretation given is that there was some difference between the studies which resulted in reappraisal working in one but not the other. However, another interpretation cannot be ruled out, which is that reappraisal was not actually effective in either study, meaning that the Milan result was due to chance.

Dismissing the results of the Milan study would support the view that increased ecological validity nullifies the effect of reappraisal in general. The OU study could have supported the effects of reappraisal in the Milan study, but has instead become a contrast to it. So, confidence in the results of the Milan study is not as strong due to the failure to replicate them in the OU study. The main motivation for testing reappraisal in this thesis was to do so while also improving ecological validity. On this view, the decision-making process in complex trading tasks like the two-index game negates the effect of reappraisal.

A related interpretation is that reappraisal may not be powerful enough to exert a measurable effect on decision making when the environment decisions are made in a more complex situation (i.e. an effect of reappraisal is small compared to the noise in the data). So, reappraisal could theoretically have an effect, but would not be practically useful for investors. It is likely that the size of the effect of reappraisal changes with each experimental setup used and the characteristics of the sample used. As before,

further research into typical effect sizes on trading biases when reappraisal is used would be valuable.

9.3.2.3 Measuring effect on emotions

The OU study found no effect of reappraisal on emotions. However, it appears that reappraisal had no effect at all so little can be drawn from the potential effect of reappraisal if it could be made effective.

A similar study could investigate alternative methods of measuring emotions. A limitation of the OU study, beyond the null effect of reappraisal, was that emotions were measured indirectly. The PANAS and other questionnaires were self-report, based on memory (albeit only slightly in the past), and summative over the course of a 5 minutes game.

The protocol directed participants to respond about the emotions they experienced while making trading decisions. However, this will include some noise as participants may have struggled to separate their experience of emotions when making decisions from their experience of emotions in general when playing the game. Participants made multiple trades over the game, sometimes dozens, and the protocol required participants to generalise over all these decisions, for some 40 different emotion items. This is obviously difficult, and may have resulted in recall issues.

An improvement in measuring emotions would capture the emotions experienced at each specific decision point. This was neatly isolated by Summers and Duxbury (2012), who measured emotions at the precise moment that participants would be making trading decisions. However, the experimental design precluded this in the OU study. To retain the greater ecological validity of the game, participants had to be free to trade continuously over the course of the game. This meant that participants made many decisions over course of each play.

It would have been very impractical to stop participants playing and report the emotions they were experiencing, every time they considered selling a position. Indeed, doing so would have vitiated the realistic conditions of the game, which were specifically included

in the study to achieve greater ecological validity. So here again, we have a trade-off contrast between increasing the ecological validity of the experiment, and the precision with which variables can be measured.

Physiological measures may be an avenue to pursue, to allow greater accuracy when measuring emotion proxies, while retaining the realistic trading conditions of the two-index game. Some possibilities are skin-conductance response, heart-rate variability, and EEG (since it has greater temporal resolution than other brain imaging methods). However, this would create an additional issue from the matching of individual decisions with the emotions measured at that time. Summers and Duxbury for example, isolated one decision at a time and measured emotions at that point. Whereas participants typically made 20 decisions in each play of the two-index game. The emotion measurements would have to be matched to each decision, and of course the number and timing of decisions would vary for every participant. The complexity of the data analysis required is one reason that this approach was not pursued in this study.

Another consideration when measuring emotions for specific decisions during the game, is how to measure situations when a participant considers closing a trade, but not does so. (OU study participants were instructed to consider their emotions at any time they were considering closing trades. However, participants may have found it difficult to recall every time this happened, and generalise across them to answer.) These situations may be just as important as the times when trades are made. The disposition effect involves choosing to hold losses more frequently than holding gains. This difference may occur when considering holding losses but choosing not to, meaning this behaviour could be crucial to understanding how a disposition effect occurs.

Unfortunately, this information may be more difficult to capture. It would be easy to identify the periods relating to when participants sold gains or losses, and compare the emotions experienced in each. Researchers would only have to cross-reference the time of closing positions with the time of the emotion measurement. Unfortunately, there is

no such easy reference for when trades are considered but not made. This emotional experience is again much easier to capture when trading decisions are made in an artificial experimental design. When participants are told they must decide to trade or not at specific times, then a researcher can be confident that the emotions experienced at that point relate to making a decision. Then it is easy to analyse these emotions based on the type of position (gain / loss) and decision made (hold / sell). The lack of ability to do this, because of using a more realistic trading environment, is another compromise here associated with greater ecological validity.

9.4 IMPLICATIONS FOR PRACTITIONERS

This section discusses the implications for practitioners – people connected with the finance industry, trading and investing. This includes traders and investors themselves, as well as others in the related activities such as investment management, trading supervision, and trading platform design.

The first contribution provides robust evidence that the disposition effect is a bias which differs between investors. Since the disposition effect is detrimental to trading performance, investors can be tested for this bias and made aware of it. Since the disposition effect is stable, investors who present a disposition effect on one occasion are likely to retain it over time. If an investor has a disposition effect, this is something both worth knowing and worth addressing: investors to have a high disposition effect can attempt interventions to reduce this bias.

The second contribution allows more insight into someone's trading patterns. The disposition effect alone doesn't capture the whole story – bias could be due to behaviour towards gains, or losses, or both. When investigating an investor's trading patterns, practitioners should want to measure behaviour towards gains and losses individually, as well as the overall disposition effect. This will give a more detailed account of how they trade, how they differ from other investors, and how they may deviate from optimal trading patterns and strategies.

When investors demonstrate a bias towards gains or losses, interventions could be attempted to target behaviour specifically towards gains or losses, depending on which bias appears to be responsible for the disposition effect in that scenario. Investors will be empowered by knowing that it is specifically gains, or specifically losses, which are causing them problems when trading.

The success of the self-report scale will also interest practitioners. A scale that has a strong correlation with actual trading behaviour could be very useful in quickly identifying potential issues in a trader's psychology. Even in its current form, the scale and subscales are indicative of someone's potential trading biases. However, the scale has a lot of potential for further development, as discussed above in the implications section. If successful, this could make it an accurate diagnostic tool for trading biases without even needing access to trading records. This use could be particularly interesting to those in trading education, who are trying to build awareness of trading biases and identify investors who could benefit from their courses.

The Milan study found that cognitive reappraisal can reduce the disposition effect (by reducing holding losses) in expert investors. So, a straight-forward implication for investors is to use this de-biasing technique during their own trading. Similarly, those responsible for managing or advising investors could pass on this technique to others.

In a wider sense, the success of the Milan study suggests that it may be possible to change the disposition effect using psychological techniques such as emotion regulation. More generally, it suggests that psychological techniques may be useful to help change financial behaviour and decisions. This supports many studies in recent years in the growing field of behavioural economics and behavioural science, and many practitioners will be interested to know that this has been successfully applied to decision-making biases in trading.

However, the contrasting results of the OU study put limits on how far these psychological techniques might be applied. The Milan and OU studies combined produced

an unexpected conclusion: reappraisal is still effective in reducing trading biases when tested with greater ecological validity, but *only* when tested with experts.

This is not terrible news for de-biasing techniques to improve financial decisions. Most people who trade on financial markets should eventually gain the experience required to become experts (the reasons for why there might be differences between novices and experts are discussed earlier in this chapter). So, reappraisal should still be useful for regular and experienced investors.

However, cognitive reappraisal does not appear to be an appropriate strategy for novices who make infrequent financial decisions: using it may reduce their performance, rather than improve it. Those interested in giving advice to the wider public should not advise the adoption of cognitive reappraisal. Likewise, individuals looking to improve their financial decisions but who have little experience of financial markets would be better to not adopt cognitive reappraisal, until they have gained some expertise with the trading environment.

Novices may also benefit from being sceptical about psychological techniques aimed at changing financial decision making in general. The results of these studies suggest that psychological techniques which are effective in the lab may not be with novices, when transferred to real-world setting and the cognitive load produced by the real-world task is high. So, this gives pause for thought about the implications of the wider behavioural economics literature. It may not be so easy for the general population to apply and benefit from psychological interventions found in the lab.

9.5 OVERALL CONCLUSIONS

This thesis has been successful in demonstrating properties of the disposition effect and its constituent biases (cutting gains and holding losses). All three have been shown to be reliably measured using a sophisticated trading game in the lab, and this has been demonstrated for all three biases in three separate studies.

Convergent validity has been demonstrated between the trading-based measure of the disposition effect and a self-report scale. This supports the status of the disposition effect as a persistent feature of trading behaviour, and supports the value of further development of the self-report scale. One major limitation of this work is the extent to which decision-making differs in short-term versus long-term trading situation; however, convergent validity with the self-report scale partly mitigates this uncertainty.

Convergent validity has also been demonstrated convincingly for holding losses between the trading game and the scale. There is some evidence that convergent validity for cutting gains could be achieved with further development of the self-report scale.

There is good evidence that cutting gains and holding losses are independent biases, this being demonstrated in both the trading game and the scale independently. This reveals that the disposition effect is not actually a unitary bias at all, but it is the sum effect of these two independent trading behaviours.

Having made these contributions about the nature and measurement of the disposition effect and its constituent biases, this thesis investigated the effect of reappraisal while improving external validity when in testing it. It also investigated possible mechanisms for the effect of reappraisal in the emotions experienced while trading.

One study using experts (retail investors) found that reappraisal was still effective with improvements in external validity. Although issues with the dataset from this study limit the confidence that can be placed on those findings, if true the findings from this study suggest reappraisal could be a simple yet powerful technique for improving decision making in trading.

A second study with novices failed to find an effect of reappraisal on trading behaviour, or on the emotions experienced during trading. Given that reappraisal failed to influence trading behaviour, no strong conclusions can be drawn about the effect of reappraisal on emotions, nor about the mechanism that may produce reappraisal's effect.

There are two main interpretations of the result of the second study: either the first study's result was spurious, or some difference in novices (as opposed to experts) nullified the effect of reappraisal. Based on the latter type of interpretation, further research could examine: differences in the process of decision-making between experts and novices, the cognitive load placed on experts and novices, and whether other types of cognitive reappraisal or emotion regulation are more effective as de-biasing strategies for novices.

The results of this thesis for the nature of trading biases are very encouraging, in establishing these biases as stable determinants of trading behaviour that can be reproduced in the lab. The results for the effect of cognitive reappraisal are more complicated. It was effective for experts in ecologically valid experimental settings, but no longer effective for novices.

Cognitive reappraisal has the potential to be a simple method to help the public improve their financial decisions, by giving them just a few sentences of instructions about changing their mind-set. As noted by Odean and others, investors with a disposition can incur substantial reductions in returns. However, using cognitive reappraisal to improve decisions this has turned out to be more difficult to achieve in practice. The failure to replicate the effect here in novices is a reminder that ecological validity is paramount. When we want to generalise promising results from the lab to the field, demonstrating these results again with increased ecological validity is an essential test of their robustness in the real world.

REFERENCES

- Allison, P. D. (2001). *Missing data* (Vol. 136): Sage publications.
- Barber, B. M., Lee, Y. T., Liu, Y. J., & Odean, T. (2007). Is the aggregate investor reluctant to realise losses? Evidence from Taiwan. *European Financial Management*, *13*(3), 423-447.
- Barberis, N., & Xiong, W. (2009). What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation. *The Journal of Finance*, *64*(2), 751-784.
- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in cognitive science*, *6*(6), 248-254.
- Baumann, N., Kaschel, R., & Kuhl, J. (2005). Striving for Unwanted Goals: Stress-Dependent Discrepancies Between Explicit and Implicit Achievement Motives Reduce Subjective Well-Being and Increase Psychosomatic Symptoms. *Journal of personality and social psychology*, *89*(5), 781.
- Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex* 1. *Cognition*, *50*(1-3), 7-15.
- Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review*, *91*(1), 79-98.
- Bodenhausen, G. V., Kramer, G. P., & Süsser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of personality and social psychology*, *66*(4), 621-632.
- Brown, A. L., & Kagel, J. H. (2009). Behavior in a simplified stock market: the status quo bias, the disposition effect and the ostrich effect. *Annals of Finance*, *5*(1), 1-14.
- Brown, P., Chappel, N., Da Silva Rosa, R., & Walter, T. (2006). The Reach of the Disposition Effect: Large Sample Evidence Across Investor Classes*. *International Review of Finance*, *6*(1-2), 43-78.
- Brunswik, E. (1956). Perception and the representative design of experiments. Berkeley: CA: University of California Press.
- Bryman, A., & Bell, E. (2015). *Business research methods*: Oxford University Press, USA.
- Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of personality and social psychology*, *67*(3), 366.
- Camerer, C., & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives*, *9*(2), 209-219.
- Camerer, C. F., & Johnson, E. J. (1991). 8 The process-performance paradox in expert judgment. *Toward a general theory of expertise: Prospects and limits*, 195.

- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2), 81.
- Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, 280(5364), 747.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research*, 1(2), 245-276.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive psychology*, 4(1), 55-81.
- Chui, P. M. (2001). An experimental study of the disposition effect: Evidence from Macau. *The journal of psychology and financial markets*, 2(4), 216-222.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*: Lawrence Erlbaum.
- Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment*, 4(1), 5.
- Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics. G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart*, 97-118.
- Dacey, R., & Zielonka, P. (2008). A detailed prospect theory explanation of the disposition effect. *The Journal of Behavioral Finance*, 9(1), 43-50.
- Damasio, A. R. (1994). *Descartes' Error: Emotion*: GP Putnam.
- Davies, J. J. (1968). On the scientific method: how scientists work.
- De Bondt, W. F. M., & Thaler, R. H. (1990). Do security analysts overreact? *The American Economic Review*, 80(2), 52-57.
- Derakshan, N., Eysenck, M. W., & Myers, L. B. (2007). Emotional information processing in repressors: The vigilance-avoidance theory. *Cognition & Emotion*, 21(8), 1585-1614.
- Deutsch, R. M. (1977). The New Nuts Among the Berries. How Nutrition Nonsense Captured America. *Bull Publishing Co., Palo Alto*, 281-292.
- Dhar, R., & Wertenbroch, K. (2000). Consumer choice between hedonic and utilitarian goods. *Journal of Marketing Research*, 37(1), 60-71.
- Dhar, R., & Zhu, N. (2006). Up close and personal: Investor sophistication and the disposition effect. *Management Science*, 52(5), 726.
- Epstein, S. (2010). Demystifying Intuition: What It Is, What It Does, and How It Does It. *Psychological Inquiry*, 21(4), 295-312.
- Epstein, S., Lipson, A., Holstein, C., & Huh, E. (1992). Irrational reactions to negative outcomes: Evidence for two conceptual systems. *Journal of personality and social psychology*, 62(2), 328-339.

- Erber, R., Wegner, D. M., & Therriault, N. (1996). On being cool and collected: Mood regulation in anticipation of social interaction. *Journal of personality and social psychology, 70*(4), 757.
- Ericsson, K., & Lehmann, A. (1996). Expert and exceptional performance: evidence of maximal adaption to task constraints. *Annual review of psychology, 47*.
- Esch, T., Fricchione, G. L., & Stefano, G. B. (2003). The therapeutic use of the relaxation response in stress-related diseases. *Signature, 9*(2), 34.
- Evans, J. S. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10), 454-459.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383-417.
- Feldman, M. S., & March, J. G. (1981). Information in organizations as signal and symbol. *Administrative Science Quarterly, 26*(2), 171-186.
- Feng, L., & Seasholes, M. S. (2005). Do investor sophistication and trading experience eliminate behavioral biases in financial markets? *Review of Finance, 9*(3), 305-351.
- Fenton-O'Creevy, M., Soane, E., Nicholson, N., & Willman, P. (2011). Thinking, feeling and deciding: The influence of emotions on the decision making and performance of traders. *Journal of Organizational Behavior, 32*(8), 1044-1061.
- Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13*(1), 1-17.
- Friedman, M. (1953). The methodology of positive economics. *The Philosophy of economics: an anthology, 2*, 180-213.
- Ganzach, Y. (2000). Judging Risk and Return of Financial Assets* 1. *Organizational behavior and human decision processes, 83*(2), 353-370.
- Gary, M. S., Dosi, G., & Lovallo, D. (2007). Boom and Bust Behavior: On the Persistence of Strategic Decision Biases and their Collective Outcome. *LEM Papers Series*.
- Genesove, D., & Mayer, C. (2001). Loss aversion and seller behavior: Evidence from the housing market. *The quarterly journal of economics, 116*(4), 1233-1260.
- Glick, I. O. (1957). *A social psychological study of futures trading*. University of Chicago, Department of Sociology.
- Grinblatt, M., & Keloharju, M. (2001). What makes investors trade? *The Journal of Finance, 56*(2), 589-616.
- Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General psychology, 2*(3), 271.
- Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology, 39*(3), 281-291.

- Gross, L. (1982). *The art of selling intangibles: How to make your million (\$) by investing other people's money*: Prentice Hall.
- Hafenbrack, A. C., Kinias, Z., & Barsade, S. G. (2014). Debiasing the mind through meditation mindfulness and the sunk-cost bias. *Psychological Science, 25*(2), 369-376.
- Hammond, K. R. (1998). Representative design. Retrieved October, 9, 1998.
- Heath, C., Huddart, S., & Lang, M. (1999). Psychological factors and stock option exercise. *Quarterly Journal of Economics, 601-627*.
- Heilman, R. M., Cri an, L. G., Houser, D., Miclea, M., & Miu, A. C. (2010). Emotion regulation and decision making under risk and uncertainty. *Emotion, 10*(2), 257.
- Heisler, J. (1994). Loss aversion in a futures market: An empirical test. *Review of Futures Markets, 13*(3), 793-826.
- Henslin, J. M. (1967). Craps and magic. *American Journal of Sociology, 73*(3), 316-330.
- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational research methods, 1*(1), 104-121.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185.
- Horton, S., Baker, J., & Schorer, J. (2008). Expertise and aging: maintaining skills through the lifespan. *European Review of Aging and Physical Activity, 5*(2), 89-96.
- Izard, C. E. (1990). Facial expressions and the regulation of emotions. *Journal of personality and social psychology, 58*(3), 487.
- Johnson, E. J., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of personality and social psychology, 45*(1), 20-31.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*: Lawrence Erlbaum Associates, Inc.
- Kadous, K., Tayler, W. B., Thayer, J. M., & Young, D. (2014). Individual characteristics and the disposition effect: the opposing effects of confidence and self-regard. *Journal of Behavioral Finance, 15*(3), 235-250.
- Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*(6), 515-526.
- Kahneman, D., Ritov, I., & Schkade, D. (1999). Economic preferences or attitude expressions: An analysis of dollar responses to public issues. *Journal of Risk and Uncertainty, 19*(1), 203-235.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society, 263-291*.

- Kahneman, D., & Tversky, A. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 91, 293-315.
- Kahneman, D., & Tversky, A. (2000). Choices, values, and frames: Cambridge University Press.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*.
- Kanfer, R., Ackerman, P. L., Murtha, T. C., Dugdale, B., & Nelson, L. (1994). Goal setting, conditions of practice, and task performance: A resource allocation perspective. *Journal of Applied Psychology*, 79(6), 826.
- Kassin, S. M. (2003). Essentials of psychology: Prentice Hall.
- Kaustia, M. (2010). Prospect theory and the disposition effect. *Journal of Financial and Quantitative Analysis*, 45(03), 791-812.
- Kelley, H. H. (1973). The Processes of Causal Attribution¹. *American Psychologist*, 107.
- Keren, G., & Teigen, K. H. (2005). Yet another look at the heuristics and biases approach. *Handbook of judgment and decision making*, 91-111.
- Knight, F. H. (1921). Risk, uncertainty and profit: Houghton Mifflin, Boston.
- Koole, S. L. (2009). The psychology of emotion regulation: An integrative review. *Cognition & Emotion*, 23(1), 4-41.
- Laplace, P. (1951). A philosophical essay on probability [English translation, original work published 1814: Essai philosophique sur les probabilités]: New York: Dover Publications.
- Lee, H. J., Park, J., Lee, J. Y., & Wyer Jr, R. S. (2008). Disposition effects and underlying mechanisms in e-trading of stocks. *Journal of Marketing Research*, 45(3), 362-378.
- Lee, J.-S., Shen, C.-H., & Yen, P.-H. (2010). The study of the fund investors' disposition effect vis-à-vis herding redemption and non-herding redemption: evidence from quantile regression least square dummy variable estimator. *Journal of Management and System*, 17(1).
- Lehenkari, M. (2012). In search of the underlying mechanism of the disposition effect. *Journal of Behavioral Decision Making*, 25(2), 196-209.
- Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of personality and social psychology*, 81(1), 146.
- Lo, A. W., & Repin, D. V. (2002). The psychophysiology of real-time financial risk processing. *Journal of cognitive neuroscience*, 14(3), 323-339.
- Locke, P. R., & Onayev, Z. (2005). Trade duration: Information and trade disposition. *Financial Review*, 40(1), 113-129.
- Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior research methods*, 38(1), 88-91.

- MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological review*, *55*(2), 95.
- Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition & Emotion*, *23*(2), 209-237.
- Mayer, J. D., Gaschke, Y. N., Braverman, D. L., & Evans, T. W. (1992). Mood-congruent judgment is a general effect. *Journal of personality and social psychology*, *63*(1), 119-132.
- McClure, S. M., Botvinick, M. M., Yeung, N., Greene, J. D., & Cohen, J. D. (2007). Conflict monitoring in cognition-emotion competition. *Handbook of emotion regulation*, 204-228.
- McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*(5695), 503-507.
- Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of economic theory*, *3*(4), 373-413.
- Miu, A. C., & Crişan, L. G. (2011). Cognitive reappraisal reduces the susceptibility to the framing effect in economic decision making. *Personality and Individual Differences*, *51*(4), 478-482.
- Niedenthal, P. M. (2007). Embodying emotion. *Science*, *316*(5827), 1002.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*: Prentice-Hall Englewood Cliffs, NJ.
- Nunnally, J. C., & Bernstein, I. (1994). *Psychological theory*: New York: McGraw-Hill.
- Odean, T. (1998). Are investors reluctant to realize their losses? *Journal of Finance*, 1775-1798.
- Outhwaite, W. (1987). Realism and social science *New Philosophies of Social Science* (pp. 45-60): Springer.
- Pedhazur, E. J., & Schmelkin, L. P. (2013). *Measurement, design, and analysis: An integrated approach*: Psychology Press.
- Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology*, *74*(2), 116-143.
- Rau, H. A. (2014). The disposition effect and loss aversion: Do gender differences matter? *Economics Letters*, *123*(1), 33-36.
- Rau, H. A. (2015). The disposition effect in team investment decisions: Experimental evidence. *Journal of Banking & Finance*, *61*, 272-282.
- Richards, D. (2012). *The disposition effect, dual process theory and emotion regulation*. Open University.
- Rosen, M., Salas, E., Lyons, R., & Fiore, S. (2008). Expertise and naturalistic decision making in organizations: Mechanisms of effective decision making. *The Oxford handbook*

of organizational decision making: Psychological and management perspectives, 211-230.

Rothermund, K., Voss, A., & Wentura, D. (2008). Counter-regulation in affective attentional biases: A basic mechanism that warrants flexibility in emotion and motivation. *Emotion*, 8(1), 34.

Rottenstreich, Y., & Hsee, C. K. (2001). Money, kisses, and electric shocks: On the affective psychology of risk. *Psychological Science*, 12(3), 185-190.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300(5626), 1755.

Schlarbaum, G. G., Lewellen, W. G., & Lease, R. C. (1978). Realized returns on common stock investments: The experience of individual investors. *Journal of Business*, 299-325.

Schoenemann, P. T., Sheehan, M. J., & Glotzer, L. D. (2005). Prefrontal white matter volume is disproportionately larger in humans than in other primates. *Nature Neuroscience*, 8(2), 242-252.

Schunk, D., & Betsch, C., (2006). Explaining heterogeneity in utility functions by individual differences in decision modes. *Journal of Economic Psychology*, 27(3), 386-401.

Schwarz, N. (1990). *Feelings as information: Informational and motivational functions of affective states*: Zentrum für Umfragen, Methoden u. Analysen eV (ZUMA).

Seru, A., Shumway, T., & Stoffman, N. (2010). Learning by trading. *Review of Financial studies*, 23(2), 705-739.

Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance*, 40(3), 777-790.

Shiv, B., & Fedorikhin, A. (1999). Heart and mind in conflict: The interplay of affect and cognition in consumer decision making. *Journal of Consumer Research*, 26(3), 278-292.

Shiv, B., Loewenstein, G., Bechara, A., Damasio, H., & Damasio, A. R. (2005). Investment behavior and the negative side of emotion. *Psychological Science*, 16(6), 435.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological methods*, 7(4), 422.

Shu, P. (2005). The research of the disposition effect in Taiwan stock market. *Taiwan Banking and Finance Quarterly*, 6, 21-33.

Siemer, M., & Reisenzein, R. (1998). Effects of mood on evaluative judgements: Influence of reduced processing capacity and mood salience. *Cognition & Emotion*, 12(6), 783-805.

Slooman, S. (2002). TWO SYSTEMS OF REASONING. T. GILOVICH, D. GRIFFIN & D. Kahneman (Eds.). *Heuristics and biases*: New York: Cambridge University Press.

- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, 177(3), 1333-1352.
- Sokol-Hessner, P., Hsu, M., Curley, N. G., Delgado, M. R., Camerer, C. F., & Phelps, E. A. (2009). Thinking like a trader selectively reduces individuals' loss aversion. *Proceedings of the National Academy of Sciences*, 106(13), 5035.
- Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, 18(3), 275-290.
- Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of personality and social psychology*, 94(4), 672.
- Stevens, J. (1996). Exploratory and confirmatory factor analysis. *Applied multivariate statistics for the social sciences*, 362-428.
- Summers, B., & Duxbury, D. (2012). Decision-dependent emotions and behavioral anomalies. *Organizational behavior and human decision processes*.
- Tamir, M., Chiu, C. Y., & Gross, J. J. (2007). Business or pleasure? Utilitarian versus hedonic considerations in emotion regulation. *Emotion*, 7(3), 546.
- Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological bulletin*, 116, 21-21.
- Thaler, R. H. (1999). The end of behavioral finance. *Financial Analysts Journal*, 12-17.
- Thaler, R. H., & Johnson, E. J. (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science*, 643-660.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), 207-232.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297-323.
- Tversky, A. K., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4754), 1124-1131.
- Wallace, J. C., Edwards, B. D., Shull, A., & Finch, D. M. (2009). Examining the consequences in the tendency to suppress and reappraise emotions on task-related job performance. *Human Performance*, 22(1), 23-43.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Psychology*, 12(3), 129-140.
- Watson, D., & Clark, L. A. (1999). The PANAS-X: Manual for the positive and negative affect schedule-expanded form.

- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, 54(6), 1063.
- Weber, M., & Camerer, C. F. (1998). The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior & Organization*, 33(2), 167-184.
- Weber, M., & Welfens, F. (2007). *An individual level analysis of the disposition effect: Empirical and experimental evidence*: Univ., Sonderforschungsbereich 504.
- Werner, N. S., Jung, K., Duschek, S., & Schandry, R. (2009). Enhanced cardiac perception is associated with benefits in decision making. *Psychophysiology*, 46(6), 1123-1129.
- Yamagishi, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S., & Cook, K. S. (2009). The private rejection of unfair offers and emotional commitment. *Proceedings of the National Academy of Sciences*, 106(28), 11520-11523.
- Yee, J., & Lins, J. (2011). Computational models for the two index game: xDelia working paper.
- Yee, J., & Lins, J. (2011). Disposition effect among professional traders and private investors Saxo Bank.
- Zajonc, R. B. (1998). Emotions.
- Zeelenberg, M., Van Dijk, W. W., & Manstead, A. S. (1998a). Reconsidering the relation between regret and responsibility. *Organizational behavior and human decision processes*, 74(3), 254-272.
- Zeelenberg, M., Van Dijk, W. W., SR Manstead, A., & der Pligt, J. (1998b). The experience of regret and disappointment. *Cognition & Emotion*, 12(2), 221-230.

APPENDIX 1: A MATHEMATICAL EXPLANATION OF THE DISPOSITION USING PROSPECT THEORY

1. The value function $v(x)$ for a gain or loss position is not a linear function of x , but is transformed for gains using the equation $v(x) = x^\alpha$ (or for losses $v(x) = -\lambda(-x)^\beta$). α (and β) relates to the diminishing marginal effect of increases (or decreases) in wealth on expected utility as the price moves away from the reference price (i.e. the purchase price), and takes values between 0 and 1. λ refers to loss aversion – the preference for a loss over a gain of equal value, which will be referred to later. Both α and β have been estimated at 0.88 and λ at 2.25 (Tversky and Kahneman 1992).
2. When making decisions about whether to buy or sell, the external variables considered are: x and its transformation into $v(x)$; h , a further movement from this position that will happen if the stock is held, and which could be positive or negative; p , the probability of h being positive, and $1 - p$, the probability of h being negative.
3. In the basic explanation, h is considered equal to or less than x , such that the investor will not switch from a gain position to a loss position, and vice versa; at best, they can get back to break-even. The only possibilities are an increase of h or a decrease of h .
4. For the simplified model explanation, p and $1 - p$ are both fixed at 0.5.
5. The value function produces asymmetric changes in expected utility when a movement of h or $-h$ is considered on top of an existing gain or loss position of x or $-x$. For gains ($x > 0$) the value function is concave due to operation of α , such that the first derivative is always positive and the second derivative is always negative. Thus, as x increases, a positive movement of h will always produce a magnitude of change in $v(x)$ smaller than a negative movement of h . The average of these values will be lower than the expected utility with no movement, i.e. selling when the price is x , where $v(x) = x^\alpha$. So, given a 0.5 probability of h and 0.5 probability of $-h$, expected utility when gambling is negative compared to immediately selling. Selling will be preferred to holding when:

$$x^\alpha > 0.5((x+h)^\alpha - x^\alpha) + 0.5(x^\alpha - (x-h)^\alpha)$$

$$\rightarrow 2x^{\alpha} > (x+h)^{\alpha} + (x-h)^{\alpha}$$

which *ceteris paribus* will always be true, thus the stock will be sold.

For losses ($x < 0$) we see the reverse: β causes the value function to be convex, such that first derivative is always positive and the second derivative is always negative. The basic shape of the curve is the same of that for gains if it were rotated 180° around the origin. An increase of h will always produce a greater positive change in $v(x)$ than a decrease of h . Thus, if there is an equal likelihood of an increase or decrease, the expected utility of holding is greater than selling immediately. Selling will be preferred when:

$$-\lambda 2(-x)^{\beta} > -0.5\lambda((-x+h)^{\beta} + (-x-h)^{\beta})$$

which *ceteris paribus* will never be true, thus the stock will be held.

To show the effect of the value function without a graph, we can differentiate.

The value function for gains is $v(x) = x^{\alpha}$, where $0 < \alpha < 1$. Thus $dv(x) / dx = \alpha x^{\alpha-1}$ and will always be positive, but $d^2v(x) / dx^2 = \alpha(\alpha-1)x^{\alpha-2}$ thus will always be negative, so the slope starts out positively and then flattens out as it tends to a 0 gradient. For losses, we have $v(x) = -\lambda(-x)^{\beta}$, $dv(x) / dx = -\lambda\beta(-x)^{\beta-1}$, $d^2v(x) / dx^2 = -\lambda\beta(\beta-1)(-x)^{\beta-2}$, which are effectively the same as the function for gains except that they begin with a minus sign so the first derivative is always negative and the second derivative is always positive.

6. The above model only takes situations where the probabilities of a gain or loss are both 0.5. Dacey and Zielonka (2008) develop this model further by allowing the probability of a gain or loss to vary, though they keep the binary possible outcomes as a positive or negative movement of h . With this revision, if the probability of a positive movement equals p and a negative movement equals $1 - p$, then selling will be preferred to holding in a gain position if:

$$x^{\alpha} > p(x+h)^{\alpha} + (1-p)(x-h)^{\alpha}$$

They demonstrated that this still produces a disposition effect. This extension still only considers movements of h which do not cause the mental account to change from gain to loss or vice versa, and where there are only two outcomes, either a positive or negative movement of h .

APPENDIX 2: RISK AVERSION AND LOSS AVERSION DEFINED MATHEMATICALLY

Loss-framed risk propensity and gain-framed risk-aversion are the results of a systematic difference between losses and gains in the expected utility of holding or selling. This difference is caused by parameters α and β in the value function of prospect theory, which in turn lead to decisions which appear to involve changes in the appetite for risk. For example, given a choice between £100 for certain (analogous to selling a gain), and a 0.5 probability each of £200 or £0, they will usually opt for the former. The expected values of these options are the same while they differ in the risk involved; however, the expected utility according to prospect theory is higher for the certain gain. For a certain loss of £100 or a 0.5 probability of losing £200 people will choose the reverse, taking the risky option over the certainty of a £100 loss. Again, the expected values of the options are the same and differ only in risk involved, but the expected utility is higher for the risky option. Investors are sometimes said to be “averse” to taking the certain loss, and willing to take on additional risk for the chance of avoiding it.

In contrast the greater weighting of losses versus gains, defined by loss aversion in the main text), is represented in the value function by the parameter λ . This has been estimated to have a value of 2.25. That is, after the expected value has been transformed into expected utility using x^α or $(-x)^\beta$, the provisional expected utility for a loss is to be multiplied by 2.25 to arrive at its final expected utility. Put another way, a transformed gain needs to be 2.25 times the transformed loss for them to be considered equal. Thus, most people refuse a bet of a 0.5 probability of a £110 gain and a 0.5 probability of a £100 loss, despite the expected value clearly being positive.

In summary, reference to diminishing marginal utility is unnecessary to explain why investors refuse a gamble of positive expected value (such as a 50/50 chance for £110 or -£100. This can be explained using only this loss aversion, where the value function is $v(x) = x$ for gains and $v(x) = \lambda x$ for losses, and $\lambda > 1$. Conversely the disposition effect can be explained by risk appetite and the stock position being framed with reference to the purchase price, and does not need to refer to loss aversion.

APPENDIX 3: REAPPRAISAL INSTRUCTION FROM THE MILAN STUDY

The reappraisal group received the following instruction prior to the last play of the simulation:

"You will now play the Two Index Game for the final time. Prior research into the psychology of trading has found that mentally distancing yourself from decisions can sometimes improve them. Now that you have played the Two Index Game several times and become familiar with how it works, it should be possible for you to try a strategy to help improve your performance.

When playing the Two Index Game this time, please imagine yourself in the role of an investment manager who is trading on behalf of a client. Please still aim to make as much money as possible, this time for your client."

This was followed by a manipulation check, to ensure that participants had in fact read this instruction, and not simply tried to click through to the next page. They were asked "How will you play the Two Index Game this time?", followed by checkboxes for "As myself" and "As an investment manager on behalf of a client".

After the manipulation page, a final page read:

"Thank you for completing the survey! You will now play the Two Index Game for the final time. As with your previous attempts at the game, please try to make as much money as possible, this time for your client. Please click "done" and return to the "How to participate" email you have been sent. Continue to the next section and play the Two Index Game for the 4th time, taking on the role of an investment manager working for a client, as described on the previous page."

The control group had no reappraisal instruction, and the final page read:

"Thank you for completing the survey! You will now play the Two Index Game for the final time. As with your previous attempts at the game, please try to make as much money as possible. Please click "done" and return to the "How to participate document" email you have been sent."

APPENDIX 4: PARTICIPANT BOOKLET FROM THE OU STUDY (FOR REAPPRAISAL GROUP PARTICIPANTS)

ID NUMBER:

Thank you again for participating in this study today. This booklet will tell you how to complete each part of the study, and also includes the questionnaires to answer. Please feel free to ask the researcher if you have any questions.

OUTLINE

As a reminder, the study outline is as follows, with approximate times given:

#	Study component	Approx. minutes
	Introduction and consent form	4
1	Questionnaire 1	2
2	Two-index Game Tutorial and 1 st play of the game	14
3	2 nd play of the game	5
4	Questionnaire 2	5
5	3 rd play of the game and questionnaire 3	6
6	4 th play the game and questionnaire 4	7
7	Debrief	2
	Total	45

QUESTIONNAIRE 1

This scale consists of a number of words and phrases that describe different feelings and emotions. Read each item and then circle the appropriate answer.

PLEASE INDICATE TO WHAT EXTENT YOU FEEL THIS WAY **RIGHT NOW** (this is, at the present moment)

Please answer quickly with your first impression – the questionnaire should only take a minute or two.

	1	2	3	4	5
	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
1. Active	1	2	3	4	5
2. Disgusted	1	2	3	4	5
3. Calm	1	2	3	4	5
4. Guilty	1	2	3	4	5
5. Enthusiastic	1	2	3	4	5
6. Attentive	1	2	3	4	5
7. Afraid	1	2	3	4	5
8. Nervous	1	2	3	4	5
9. Distressed	1	2	3	4	5
10. Shaky	1	2	3	4	5
11. Excited	1	2	3	4	5
12. Determined	1	2	3	4	5

	1	2	3	4	5
	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
13. Strong	1	2	3	4	5
14. Hostile	1	2	3	4	5
15. Frightened	1	2	3	4	5
16. Scornful	1	2	3	4	5
17. Proud	1	2	3	4	5
18. Relaxed	1	2	3	4	5
19. Alert	1	2	3	4	5
20. Jittery	1	2	3	4	5
21. Interested	1	2	3	4	5
22. Irritable	1	2	3	4	5
23. Upset	1	2	3	4	5
24. Loathing	1	2	3	4	5
25. Angry	1	2	3	4	5
26. Ashamed	1	2	3	4	5
27. Inspired	1	2	3	4	5
28. At ease	1	2	3	4	5
29. Scared	1	2	3	4	5
30. Concentrating	1	2	3	4	5

TWO-INDEX GAME TUTORIAL AND 1ST PLAY

Entering your player name in the game

When asked for **Player Name** at the start of the game, please use the **ID** you have been given by the researcher.

On each play of the game, please enter your ID with a suffix as follows:

- for the tutorial and first play, Player Name = xxx-1
- for the second play, Player Name = xxx-2
- for the third play, Player Name = xxx-3
- for the fourth play, Player Name = xxx-4

Tutorial

As well as the tutorial, there are written notes to help you understand how the game works. You may wish to read these after the tutorial, before continuing to the first practice game. You may also refer to these notes during the study to remind yourself of anything you have forgotten.

Open the folder named “Tutorial and Play 1” on the desktop, and double click on “TwoIndexGame”.

When asked for your **Player Name**, enter your **ID** followed by “-1”, so:

xxx-1.

When asked if this is the first time you have played, click “Yes” to enter the tutorial.

Note that there are a couple of short practice sessions at the end of the tutorial itself. Sometimes, a box will pop up telling you the level has changed. Simply click the button “Continue Playing” to continue.

During the tutorial, text will often pop up accompanied by an “Ok” button. **Before carrying out any action** that the text prompts you to do, **please click the “Ok” button first**. This helps the game to not crash! (Otherwise you will need to start the tutorial again). Click “Ok” on the welcome pop-up box then click “Start Game” in the bottom right corner to begin the tutorial.

Once the tutorial is over, a pop-up box says ‘You have completed the tutorial. Please click “Ok”, but **do not click “start game”** yet.

1st play of the game

You now may wish to read the additional notes written about how to play the Two-Index Game, before continuing to the first practice game. However, if you feel confident in how the game works, and particularly in how the “sell” buttons work, there is no need to.

If you need a break for the bathroom, please do so either now, or in the gaps following the Questionnaires, rather than during the game.

When you are ready press “Start Game” again in the bottom right corner. This will begin your first full play of the game, and your first full practice, which will last 5 minutes in total. There are four levels in this practice.

Each time a pop-up box tells you the level has changed, simply click “continue playing” to continue.

2nd play of the game

This is your second practice, and **will last for 5 minutes**. This 2nd play only involves level 4 of the game, meaning you can use all the buttons from the beginning of the game.

Open the folder on the desktop named “Play 2”, and double click on “TwoIndexGame”.

For your **Player Name**, enter your **ID** followed by “-2”, so: xxx-2.

When asked if this is the first time you have played, click “No” to skip the tutorial.

At the end of the game, please close any positions you are holding before the time remaining runs out and the game ends (this simplifies data analysis).

Click “Start Game” in the bottom right corner to begin the game.

QUESTIONNAIRE 2

Each of the following problems presents a choice between two options. Each problem is presented with a scale ranging from 1 (representing one option) through 6 (representing the other option). For each item, please circle the number on the scale that best reflects your relative preference between the two options. Please answer quickly with your first impression.

Problem 1

You have paid a non-returnable deposit on a gold ring for someone special. It costs £200 and you have already paid £100 on it, so you owe another £100. One day, you see in the paper that a new jewellery store is selling the same ring for only £90 as a special sale. The new store is across the street from the old one. If you decide to get the ring from the new store, you will not be able to get your money back from the old store, but you would save £10 overall.

Would you be more likely to continue paying at the old store or buy from the new store?

1 2 3 4 5 6

Most likely to
continue paying at the old store

Most likely to
buy from the new store

Problem 2

You enjoy playing tennis, but you really love crown green bowls. You just became a member of a tennis club, and of a bowls club, both at the same time. The membership to your tennis club costs £200 per year and the membership to your bowls club £50 per year. During the first week of both memberships, you develop an elbow injury. It is painful to play either tennis or bowling. Your doctor tells you that the pain will continue for about a year.

Would you be more likely to play tennis or bowling in the next six months?

1 2 3 4 5 6

Most likely to
play tennis

Most likely to
play bowling

Problem 3

You have been looking forward to this year's Halloween party. You have the right cape, the right wig, and the right hat. All week, you have been trying to perfect the outfit by cutting out a large number of tiny stars to glue to the cape and the hat, and you still need to glue them on. On the day of Halloween, you decide that the outfit looks better without all these stars you have worked so hard on.

1 2 3 4 5 6

Most likely to
wear stars

Most likely to
not wear stars

Problem 4

After a large meal at a restaurant, you order a big dessert with chocolate and ice cream. After a few bites you find you are full and you would rather not eat any more of it.

Would you be more likely to eat more or to stop eating it?

1 2 3 4 5 6

Most likely to
eat more

Most likely to
stop eating

Problem 5

You are in a hotel room for one night and you have paid £6.95 to watch a film on pay TV. Then you discover that there is a film you would much rather like to see on one of the free TV channels. You only have time to watch one of the two films.

Would you be more likely to watch the film on pay TV or on the free TV channel?

1 2 3 4 5 6

Most likely to
watch pay TV

Most likely to
watch free TV

Problem 6

You have been asked to give a toast at your friend's wedding. You have worked for hours on a story about you and your friend taking cooking lessons, but you still have some work to do on it. Then you realize that you could finish writing the speech faster if you start over and tell the funnier story about the dance lessons you took together.

Would you be more likely to finish the toast about cooking or rewrite it to be about dancing?

1 2 3 4 5 6

Most likely to
write about cooking

Most likely to
write about dancing

Problem 7

You decide to learn to play a musical instrument. After you buy an expensive cello, you find you are no longer interested. Your neighbour is moving and you are excited that she is leaving you her old guitar, for free. You'd like to learn how to play it.

Would you be more likely to practice the cello or the guitar?

1 2 3 4 5 6

Most likely to
play cello

Most likely to
play guitar

Problem 8

You and your friend are at a the cinema together. Both you and your friend are getting bored with the storyline. You'd hate to waste the money spent on the ticket, but you both feel that you would have a better time at the coffee shop next door. You could sneak out without other people noticing.

Would you be more likely to stay or to leave?

1 2 3 4 5 6

Most likely to
stay

Most likely to
leave

Problem 9

You and your friend have driven halfway to a resort. Both you and your friend feel sick. You both feel that you would have a much better weekend at home. Your friend says it is "too bad" you already drove halfway, because you both would much rather spend the time at home. You agree.

Would you be more likely to drive on or turn back?

1 2 3 4 5 6

Most likely to
drive on

Most likely to
turn back

Problem 10

You are painting your bedroom with a striped pattern in your favourite colours. It takes a long time to do. After you finish two of the four walls, you realize you would have preferred one solid colour instead of the striped pattern. You have enough paint left over to redo the entire room in the solid colour. It would take you the same amount of time as finishing the striped pattern on the two walls you have left.

Would you be more likely to finish the striped pattern or to redo the room in the solid colour?

1 2 3 4 5 6

Most likely to
finish striped pattern

Most likely to
redo with a solid colour

3RD PLAY OF THE GAME & QUESTIONNAIRE 3

3rd Play of the game

This is your 3rd play of the game, and **will last for 5 minutes**. You can use all the buttons from the beginning of the game.

Remember that your performance on this play will determine your chances of winning a voucher in the raffle. Your chances of winning are directly proportional to how well you do compared to other participants.

Open the folder named “Play 3”, and double click on “TwoIndexGame”.

For your **Player Name**, enter your **ID** followed by “-3”, so: xxx-3.

When asked if this is the first time you have played, click “No” to skip the tutorial.

At the end of the game, please close any positions you are holding before the time remaining runs out and the game ends (this simplifies data analysis).

Click “Start Game” in the bottom right corner to begin the game.

Remember to try to make as much money as possible. Good luck!

QUESTIONNAIRE 3

Part 1

This scale consists of a number of words and phrases that describe different feelings and emotions. Read each item and then circle the appropriate answer.

PLEASE INDICATE TO WHAT EXTENT YOU FELT THIS WAY WHEN YOU WERE MAKING DECISIONS **ABOUT WHETHER OR NOT TO CLOSE POSITIONS** IN THE GAME

Please answer quickly with your first impression.

	1	2	3	4	5
	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
1. Disappointed	1	2	3	4	5
2. Regretful	1	2	3	4	5
3. Happy	1	2	3	4	5
4. Satisfied	1	2	3	4	5
5. Proud	1	2	3	4	5

How responsible did you feel for the profits and losses which you made, when you took the decision to close positions?

1 2 3 4 5

Part 2

This scale consists of a number of words and phrases that describe different feelings and emotions. Read each item and then circle the appropriate answer.

PLEASE INDICATE TO WHAT EXTENT YOU FELT THIS WAY WHEN YOU WERE MAKING DECISIONS **ABOUT WHETHER OR NOT TO CLOSE POSITIONS** IN THE GAME. Please answer quickly with your first impression – the questionnaire should only take a minute or two.

	1	2	3	4	5
	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
1. Active	1	2	3	4	5
2. Disgusted	1	2	3	4	5
3. Calm	1	2	3	4	5
4. Guilty	1	2	3	4	5
5. Enthusiastic	1	2	3	4	5
6. Attentive	1	2	3	4	5
7. Afraid	1	2	3	4	5
8. Nervous	1	2	3	4	5
9. Distressed	1	2	3	4	5
10. Shaky	1	2	3	4	5
11. Excited	1	2	3	4	5
12. Determined	1	2	3	4	5
13. Strong	1	2	3	4	5

	1	2	3	4	5
	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
14. Hostile	1	2	3	4	5
15. Frightened	1	2	3	4	5
16. Scornful	1	2	3	4	5
17. Proud	1	2	3	4	5
18. Relaxed	1	2	3	4	5
19. Alert	1	2	3	4	5
20. Jittery	1	2	3	4	5
21. Interested	1	2	3	4	5
22. Irritable	1	2	3	4	5
23. Upset	1	2	3	4	5
24. Loathing	1	2	3	4	5
25. Angry	1	2	3	4	5
26. Ashamed	1	2	3	4	5
27. Inspired	1	2	3	4	5
28. At ease	1	2	3	4	5
29. Scared	1	2	3	4	5
30. Concentrating	1	2	3	4	5

PLEASE TURN OVER AND READ THE NEXT SECTION BEFORE PLAYING THE GAME THE FINAL TIME

4TH PLAY OF THE GAME & QUESTIONNAIRE 4

4th Play of the game

This is your last play of the game, and **will last for 5 minutes**. You can use all the buttons from the beginning of the game.

Remember that your performance on this play will determine your chances of winning a voucher in the second raffle. Your chances of winning are directly proportional to how well you do compared to other participants.

Now that you have played the Two-Index Game several times and become familiar with how it works, you can be given a strategy to try and improve your performance.

When playing the two-index game this time, please imagine you are trading for someone else. You could be an investment manager trading on behalf of a client, or a pension manager, or simply someone managing investments on behalf of a friend or family member. Try to imagine this vividly when you are making decisions.

Please tick the box to confirm that you have read the instruction in the paragraph above, about improving your performance:

Open the folder named “Play 4”, and double click on “TwoIndexGame”.

For your **Player Name**, enter your **ID** followed by “-4”, so: xxx-4.

When asked if this is the first time you have played, click “No” to skip the tutorial.

At the end of the game, please close any positions you are holding before the time remaining runs out and the game ends (this simplifies data analysis).

Click “Start Game” in the bottom right corner to begin the game.

Remember to imagine trading on behalf of someone else, though your aim is still to try to make as much money as possible. Good luck!

QUESTIONNAIRE 4

Part 1

This scale consists of a number of words and phrases that describe different feelings and emotions. Read each item and then circle the appropriate answer.

PLEASE INDICATE TO WHAT EXTENT YOU FELT THIS WAY WHEN YOU WERE MAKING DECISIONS **ABOUT WHETHER OR NOT TO CLOSE POSITIONS** IN THE GAME

Please answer quickly with your first impression.

	1	2	3	4	5
	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
1. Disappointed	1	2	3	4	5
2. Regretful	1	2	3	4	5
3. Happy	1	2	3	4	5
4. Satisfied	1	2	3	4	5
5. Proud	1	2	3	4	5

How responsible did you feel for the profits and losses which you made, when you took the decision to close positions?

1 2 3 4 5

Part 2

This scale consists of a number of words and phrases that describe different feelings and emotions. Read each item and then circle the appropriate answer.

PLEASE INDICATE TO WHAT EXTENT YOU FELT THIS WAY WHEN YOU WERE MAKING DECISIONS **ABOUT WHETHER OR NOT TO CLOSE POSITIONS** IN THE GAME

Please answer quickly with your first impression – the questionnaire should only take a minute or two.

	1	2	3	4	5
	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
1. Active	1	2	3	4	5
2. Disgusted	1	2	3	4	5
3. Calm	1	2	3	4	5
4. Guilty	1	2	3	4	5
5. Enthusiastic	1	2	3	4	5
6. Attentive	1	2	3	4	5
7. Afraid	1	2	3	4	5
8. Nervous	1	2	3	4	5
9. Distressed	1	2	3	4	5
10. Shaky	1	2	3	4	5
11. Excited	1	2	3	4	5
12. Determined	1	2	3	4	5

	1	2	3	4	5
	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
13. Strong	1	2	3	4	5
14. Hostile	1	2	3	4	5
15. Frightened	1	2	3	4	5
16. Scornful	1	2	3	4	5
17. Proud	1	2	3	4	5
18. Relaxed	1	2	3	4	5
19. Alert	1	2	3	4	5
20. Jittery	1	2	3	4	5
21. Interested	1	2	3	4	5
22. Irritable	1	2	3	4	5
23. Upset	1	2	3	4	5
24. Loathing	1	2	3	4	5
25. Angry	1	2	3	4	5
26. Ashamed	1	2	3	4	5
27. Inspired	1	2	3	4	5
28. At ease	1	2	3	4	5
29. Scared	1	2	3	4	5
30. Concentrating	1	2	3	4	5

Part 3

Finally, please give brief answers to the following questions about you and how you found the study.

1. How easy did you find it to imagine you were trading for someone else during the 4th play, following the instruction given after questionnaire 3? Please circle the best option:

1	2	3	4	5
Very easy	Easy	With some success	Hard	Very hard

OR:	I didn't use the instruction	I didn't read the instruction
-----	---------------------------------	----------------------------------

2. Who did you imagine you were trading on behalf of?
3. How did you find playing the Two Index Game in general?
4. Did you have a strategy for using the predictor index?
5. What is your best guess of how the predictor index is connected to the value index?

6. How familiar are you with the stock market? For example do you actively own your own shares?

7. How familiar are you with financial investments in general? For example, do you manage your own investments, or pension?

8. What is your profession?

9. Which age range do you belong to? (circle as appropriate)

Under 25

25-39

40-54

Over 55

10. Do you have any other comments about the study?

DEBRIEF

Many thanks for participating in this research. If you enjoyed the study, please tell your friends and colleagues! The researcher can give you some flyers to pass on.

However, until the study is complete, please don't discuss the specifics of the study with people who haven't taken part yet, for example specific instructions given during the study. This is to stop participants knowing exactly what to expect when they come to take part, and because different groups receive slightly different instructions.

This research is being carried out during Sept and October. If you would like to be emailed a summary of the findings of the research once the study is complete (in late 2014), please tick the box: