A Computational and Behavioural Analysis of Rationality in Contextual Preference Reversals

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Medical and Human Sciences

2014

George Farmer
School of Psychological Sciences
Contents

List of Figures .......................... 5
List of Tables .......................... 6
List of Abbreviations ................... 7
Abstract ................................ 8
Declaration .............................. 9
Copyright Statement .................... 10
Alternative Format Rationale ........... 11
Acknowledgements ...................... 12

1 Introduction .......................... 13
  1.1 Approaches to Rationality .......... 14
  1.2 The Attraction Effect ................ 18
  1.3 Theoretical Perspectives on the Attraction Effect .... 22
    1.3.1 A Violation of Rational Axioms .......... 22
    1.3.2 Mechanistic Accounts of the Attraction Effect ... 23
    1.3.3 Other Accounts of the Attraction Effect .... 28
  1.4 A Computationally Rational Analysis of the Attraction Effect .... 28
    1.4.1 Objective 1: Establish the Consequences of Violating Rational Axioms .......... 29
    1.4.2 Objective 2: A Computationally Rational Model .......... 30
    1.4.3 Objective 3: Derive Novel Predictions .......... 31
    1.4.4 Methods ................................ 31

2 The Attraction Effect in a Rapid Pointing Task .......... 32
  2.1 Introduction ........................ 32
  2.2 Method ................................ 34
    2.2.1 Participants ......................... 36
    2.2.2 Apparatus ........................... 36
    2.2.3 Design ............................... 36
    2.2.4 Procedure ......................... 37
  2.3 Results ................................ 38
  2.4 Discussion .......................... 40
### 3 The Attraction Effect as an Adaptive Response to Uncertainty

#### 3.1 Introduction

- 3.1.1 The Attraction Effect and Value-Maximisation

#### 3.2 A Model of the Attraction Effect as an Adaptive Response to Uncertainty in Expected Value Calculation

- 3.2.1 Estimating Expected Value With Ordinal Relations
- 3.2.2 Predicting the Attraction Effect

#### 3.3 Experiment 1

- 3.3.1 Methods
- 3.3.2 Analysis
- 3.3.3 Results
- 3.3.4 Discussion

#### 3.4 Experiment 2

- 3.4.1 Methods
- 3.4.2 Results

#### 3.5 Discussion

- 3.5.1 The Attraction Effect as an Adaptive Behaviour
- 3.5.2 Theoretical Implications
- 3.5.3 Conclusion

### 4 Why Expected Value Maximising Organisms Exhibit Preference Reversals and Risk Preferences

#### 4.1 Introduction

#### 4.2 Background

- 4.2.1 Expected Value and Computational Rationality
- 4.2.2 Preference Reversals
- 4.2.3 Risk Preference

#### 4.3 A Model of Computationally Rational Choice

#### 4.4 Results

- 4.4.1 Expected Value Maximization
- 4.4.2 Preference Reversals
- 4.4.3 Risk Preference

#### 4.5 General Discussion

### 5 The Attraction Effect is Adaptive for Gains but not Losses

#### 5.1 Introduction

#### 5.2 A Model of the Attraction Effect as an Adaptive Response to Uncertainty in Expected Value Calculation

- 5.2.1 Uncertainty in Expected Value Calculation
- 5.2.2 Information Provided by Ordinal relations
- 5.2.3 Combining Metric and Ordinal Information Leads to the Attraction Effect
- 5.2.4 Reduced Attraction Effect in the Loss Domain

#### 5.3 Experiment 1

- 5.3.1 Methods
- 5.3.2 Results

#### 5.4 Experiment 2

- 5.4.1 Methods
# List of Figures

1.1 The attraction effect .............................................. 19
1.2 Contextual preference reversal effects .......................... 21

2.1 The attraction effect .............................................. 33
2.2 Pointing task training stimuli ................................... 34
2.3 Example stimulus in descriptive condition .................... 35
2.4 Example stimulus in motor condition ............................. 35
2.5 Example participant’s training phase data ..................... 38
2.6 Attraction effect results .......................................... 39

3.1 The attraction effect .............................................. 45
3.2 Noisy perception of expected value .............................. 48
3.3 Environmental distribution of attraction effect expected values ...................................... 49
3.4 Model rationale for the attraction effect ....................... 51
3.5 Experiment 1 stimuli .............................................. 54
3.6 Experiment 1 results .............................................. 55
3.7 Experiment 2 stimuli .............................................. 57
3.8 Experiment 2 results .............................................. 58

4.1 The attraction effect .............................................. 67
4.2 Overview of computationally rational model .................... 74
4.3 Expected value of gambles against levels of model error parameters ................................ 78
4.4 Decoy positions used by Wedell (1991) ........................ 79
4.5 Reversal effects for model predictions and data from Wedell (1991) ................................ 80
4.6 Reversals minus inverse reversals against levels of model parameters ............................ 82
4.7 The effect of condition on risk preference ...................... 84
4.8 Conditional probability of low risk choice for levels model parameters .......................... 85
4.9 The effect of the ecological distribution of probability and value on the expected value of different risk preferences ...................................... 86

5.1 The attraction effect in gain and loss domains ................ 94
5.2 Overlapping expected value estimates ........................... 96
5.3 Different decoy effects in the loss and gain domains ........ 98
5.4 The attraction effect in the loss and gain domains .......... 98
5.5 Experiment 1 design .............................................. 102
5.6 Experiment 1 results .............................................. 104
5.7 Example two-alternative forced-choice data .................. 107
5.8 Experiment 2 results .............................................. 107

A.1 Measuring the attraction effect .................................. 135
List of Tables

2.1 Proportion of trials where the safe prospect was chosen. . . . . . . . . . . . 39
3.1 Expected value difference between prospect pairs. . . . . . . . . . . . . . 53
3.2 Stimuli values used in Experiment 2 . . . . . . . . . . . . . . . . . . . . . . . 57
List of Abbreviations

2AFC . . . . . Two-Alternative Forced-Choice
AAM . . . . . Associative Accumulation Model
ACT-R . . . . Adaptive Control of Thought - Rational
CBRA . . . . . Cognitively Bounded Rational Analysis
CDA . . . . . Context Dependent Advantage
DbS . . . . . Decision by Sampling
EPIC . . . . . Executive Process-Interactive Control
EV . . . . . . Expected Value
EVD . . . . . Expected Value Difference
IIA . . . . . Independence From Irrelevant Alternatives
IV . . . . . . Independent Variable
LCA . . . . . Leaky Competing Accumulators
MDFT . . . . Multi Alternative Decision Field Theory
MLBA . . . . Multi Attribute Linear Ballistic Accumulator
PRP . . . . . Psychological Refractory Period
PSE . . . . . Point of Subjective Equality
SD . . . . . . Standard Deviation
SEU . . . . . Subjective Expected Utility
Abstract

A Computational and Behavioural Analysis of Rationality in Contextual Preference Reversals

Doctor of Philosophy
University of Manchester
27 September 2014
George Farmer

The attraction effect reveals that people do not compare alternatives independently of one another. Instead, they make comparisons, such that preferences between two alternatives can be reversed by adding or removing otherwise irrelevant alternatives. This behaviour is particularly difficult for rational models of decision making to explain since such models require the independent evaluation of alternatives. As such these models describe preference reversal behaviour as irrational. This thesis examines what rational decision making should look like once a person’s cognitive bounds have been taken into account. The key finding is that contextual preference reversals like the attraction effect, far from being irrational, actually result from people making better decisions than they would if they assessed alternatives independently of one another.

The research was grouped into three objectives concerning the attraction effect and the rationality of human cognition. The first of these was to identify under what conditions people exhibit the attraction effect, and what consequences the behaviour has for the outcomes they experience. Two experiments revealed that the effect is only exhibited in choice sets where alternatives are approximately equal in value and therefore hard to tell apart. This finding also means that the potential negative consequences of exhibiting the attraction effect are very small, because it only occurs when alternatives are similar in value.

The second objective was to develop a computationally rational model of the attraction effect. Computational rationality is an approach that identifies what the optimal behaviour is given the constraints imposed by cognition, and the environment. Our model reveals why people exhibit the attraction effect. With the assumption that people cannot calculate expected value perfectly accurately, the model shows that in choices between prospects, the attraction effect actually results in decisions with a higher expected value. This is because noisy expected value estimates can be improved by taking into account the contextual information provided by the other alternatives in a choice set.

The final objective was to provide evidence for our model, and the computational rationality approach, by making a novel prediction. We conducted an experiment to test the model’s prediction that the attraction effect should be much reduced in the loss domain. We replicated existing attraction effect studies and extended them to the loss domain. The results replicated previous results in the gain domain and simultaneously revealed the novel finding that people did not exhibit the effect in the loss domain.

People exhibit the attraction effect as a result of making the best decision possible given the cognitive resources they have. Understanding decision making as computationally rational can provide deep insights into existing phenomena. The method allows us to ascertain the causal link between cognitive mechanisms, a person’s goal, and their decision making.
Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright Statement

i The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.
Alternative Format Rationale

This thesis is submitted in the alternative format. Each of the content Chapters (2 to 5) is written as standalone journal article. These are either under review (Chapters 3 and 4 at *Decision* and *Psychological Review* respectively) or are about to be submitted. In Chapters 2, 3 and 5 I am the first author. Chapter 4 is authored by Howes, Warren, Farmer, El-Deredy and Lewis. I have been involved in all aspects of preparing Chapter 4, including initial drafts and modelling. Together, these chapters form a coherent thread of research which is outlined in more detail in the Introduction. Authorship of each chapter is as follows:

Chapter 2: Farmer, El-Deredy, Howes and Warren

Chapter 3: Farmer, Warren, El-Deredy and Howes

Chapter 4: Howes, Warren, Farmer, El-Deredy and Lewis

Chapter 5: Farmer, Warren, El-Deredy and Howes
Acknowledgements

A huge thank you to my supervisors Paul Warren, Andrew Howes and Wael El-Deredy. They have been extremely generous with their considerable expertise, but most importantly have made the process massively rewarding and enjoyable.

I am also indebted to Duncan Brumby and Chris Janssen, without their advice and support, I doubt I would have embarked on a PhD in the first place.

Thank you to Shahd Majdub, Michelle Lim Ee Shian, Alexander Short, Henry Holdsworth, Danni Cernis, Jason Sayers, Heather Sinclair, Amy Morrison and Michael Clark who helped with much of the data collection.

My fellow students, officemates and housemates are now firm friends and have made the past three years in Manchester an absolute pleasure: Lizzi Lewis, Daniel Poole, Oana Lindner, Nayab Begum, Liam Mason, Jenna Roberts, Adam Lawrence, Gemma Barnacle, Sam Couth, Stacey Humphries, Caroline Lea-Carnall, Wendy Naylor, James Cousins, Lewis Dodd and Zoe Tsivos to name but a few. A special mention for Daniel Poole, who demanded one.

For indulging me in endless but fruitful conversations about my work, my thanks to Louis Passfield, Pete Johnson and Mark North. Finally, for the many wonderful Christmas breaks, extra special thanks to Sarah Passfield, Anna Passfield, Ellie Passfield and Maddy Passfield.
Chapter 1

Introduction

In 1956 Herb Simon observed that “decision making has become a natural meeting ground for psychological and economic theory” (p. 129). This meeting of theoretical perspectives perhaps motivated his rejection of the rational man of economics as a realistic model of human decision making, a rejection which has played a fundamental role in how we understand decision making to this day. Simon’s subsequent definition of bounded rationality generated entire research programmes based on the gap between normative theories of how decisions should be made and the often descriptive psychological theories of how decisions are made. Simon took the view that we simply do not have the cognitive resources to make calculations of the complexity that rational man would have to make. Our decisions are therefore, at best, loose approximations to what would be rational.

The concept of bounded rationality inspired Kahneman and Tversky’s seminal research into biases and heuristics (Kahneman, 2003). They explicitly set out to test how human decision making differed from the economic standard: “The rational-agent model was our starting point and the main source of our null hypotheses” (Kahneman, 2003, p. 1449). Their research created a literature on human decision making replete with examples of how people do not match the prescriptions of rational models. People predictably ignore or underweight base rate probabilities (base rate neglect Kahneman & Tversky, 1973), they will perceive that conjoint probabilities are more likely than single events (conjunction fallacy Tversky & Kahneman, 1983), and will seek evidence that confirms rather than tests their existing point of view (confirmation bias Wason, 1960, 1968). People can also show surprising inconsistency in their decisions, revealing reversals of initially stated preferences; a question framed as gains will lead to option a being preferred over option b, but when framed as the complementary losses will lead to option b being preferred to option a (framing effects Tversky & Kahneman, 1981). People can likewise be made to reverse a preference for one prospect over another simply by reducing the probability of both by a common factor (certainty effect Tversky & Kahneman, 1981).

One particular class of preference reversal phenomena that has had a large impact on decision making theory is the contextual preference reversal, in which the changing
composition of choice sets can alter previously stated preferences. One particularly important member of this class is the attraction effect (Huber, Payne, & Puto, 1982), which is another example of human decision making that does not appear to conform to how we know rational decisions should be made. The effect shows us that in a choice between two equally attractive alternatives, the addition of an obviously worse third alternative can bias people toward one of the original options. This reveals that people compare the alternatives with one another in deciding which they prefer. Consequently, the presence or absence of options that people never actually choose (and therefore you might think would be irrelevant) can nonetheless have an impact on which of the viable options people do choose. It can even lead people to reverse their preferences between the same two alternatives when those alternatives are presented either in the presence or absence of an ‘irrelevant’ third option.

Rational, value-maximising models of decision making (e.g., Luce, 1959) tell us that we should evaluate alternatives in a choice set independently of one another, and then choose the alternative to which we have assigned the highest value. The advantage of such models is that they provide a simple and parsimonious account of decision making. However, given that rational models do not explain phenomena like the attraction effect, it is the job of the decision sciences to develop causal models of behaviour and account for the apparent gap between the optimal and the observed.

This thesis aims to provide a deeper insight into the attraction effect, and other contextual preference reversals, by adopting a recently developed approach to studying the gap between normative and psychological theories of decision making. According to this computational rationality approach, the attraction effect may be interpretable as rational, if cognitive constraints are taken into account before asking what the optimal decision should be (Howes, Lewis, & Vera, 2009; Lewis, Howes, & Singh, 2014). From this perspective cognitive bounds do not prevent us from making rational decisions, they simply redefine what a rational decision is. We will provide compelling evidence that the attraction effect can indeed be understood as computationally rational. The following section expands on the theory behind this approach, before moving on to focus on its application to the attraction effect.

1.1 Approaches to Rationality

The normative rationality of economic man is not limited by mechanisms, it is concerned with logic and mathematical proofs, not with how a physically instantiated mechanism such as the brain would make the necessary calculations. As soon as a mechanism is required to make such calculations issues of memory, attention and other cognitive bounds (or ‘side conditions’ to use Simon’s term) come in to play which will limit its performance. This concept of bounded rationality brings to the forefront the notion that our
limited cognitive capacities impose constraints on our performance. The presence of biases and suboptimal decisions can then be interpreted as a natural side-effect of employing these limited mechanisms in service of our goals.

Simon (1992) regarded the process of achieving goals under conditions of bounded rationality as one of ‘satisficing’ rather than optimisation. Satisficing requires a person to simply achieve a level of performance that they consider good enough, but this need not be in any way the correct or optimal solution. “The study of the behaviour of an adaptive system is not a logical study of optimization but an empirical study of the side conditions that place limits on the approach to the optimum” (Simon, 1992, p. 157).

Anderson’s (1990) rational analysis method takes a different approach. From this point of view normative rationality is useful, not as a model of what the brain is doing, but as a methodological tool to determine why cognition takes the shape that it does. A rational analysis involves specifying the goal of the agent and a model of the environment they are acting in. A normatively rational account of how that goal can be achieved in that environment reveals the external pressure that is likely to have shaped the cognitive processes required to achieve the goal. It is important to note that simply identifying a rational solution to a given goal in a given environment does not necessarily mean that it is the correct interpretation of human behaviour. The formal model must still be plausible, consistent with existing knowledge and is only a hypothesis (which may prove to be wrong) about how the structure of the environment might have shaped cognitive processes (Chater & Oaksford, 2000).

The approach used in this thesis is to ask what is the computationally rational (Howes et al., 2009; Lewis et al., 2014) solution to a problem. This incorporates aspects of both rational analysis and bounded rationality. It involves identifying the cognitive constraints that exist, as in bounded rationality, and then asking what the optimal behaviour is given those constraints as well as those imposed by the environment and the goal (as in rational analysis). A final step is to specify a payoff function which the agent seeks to maximise. The purpose of the payoff function is to limit the range of possible behaviours that can perform the task. By then modelling how different cognitive constraints, provided by competing theories of architecture, impact on the strategy that optimises the reward we have a tool that allows us to distinguish between the architectural theories. Simply fitting cognitive model parameters to behavioural data does not achieve this because it does not permit us to distinguish whether behavioural data are the product of the huge array of strategies that cognitive architectures allow, or are due to hard cognitive constraints. The computational rationality approach has the advantage of the more profound causal explanation offered by rational analysis.

A rational analysis seeks to show how the environment we operate in shapes our cognitive processes, and can therefore abstract away from considering the mechanism that produces the behaviour. A computationally rational analysis does not abstract away from
the mechanism, but includes it in the consideration of how a goal should be achieved. This approach reveals the plausible strategies that are open to an agent or decision maker and hypothesizes that (assuming there is some feedback from the environment) they will settle on the strategy that performs best.

Howes et al. (2009) demonstrate the value of computational rationality in their examination of the psychological refractory period (PRP) effect. The cognitive architectures Executive Process-Interactive Control (EPIC, Kieras & Meyer, 1997) and Adaptive Control of Thought-Rational (ACT-R, Anderson, 1993) can both predict the PRP effect. In fact both architectures permit almost infinite implementations that could exhibit the effect and even permit implementations that would not exhibit the effect. So fitting an EPIC or ACT-R model to data does not tell us that the underlying architectural assumptions are correct, just that they can produce the same behaviour. If, however, we create an environment with a payoff function and feedback, and ask which strategies maximise reward, we restrict the possible strategies that each architecture can implement. Whilst each architecture can generate many strategies that would produce the behaviour, there are far fewer strategies that can exhibit the optimal behaviour. This restricted strategy space then results in differing predictions for each architecture which we can test against human performance. The model that most closely matches human performance under these conditions is most likely to based on the correct architectural assumptions. If an architecture resulted in a model that exceeded human performance, the question that naturally arises is why a sub-optimal strategy would be used in service of a goal when a tractable payoff function has been communicated to the agent. An agent with a goal of maximising reward, and that has feedback to that effect, should settle on the optimal strategy simply because otherwise it would simultaneously hold a goal of doing so, but not apply its resources to that goal.

In computationally rational explanations of cognition and behaviour, the external pressure on strategy selection comes from trying to maximise utility in an environment with feedback. In rational analysis the external pressure is the environment at large which shapes cognitive processes. It is therefore methodologically analogous but answering a different question. In a sense rational analysis is suited to explaining how macro-level elements of cognition are adapted to the environment. Computational rationality suggests that a strategy selection for any given task should lead to the optimal combination of resources given feedback and a payoff function. This can then be leveraged to determine hard cognitive constraints, such as in distinguishing between EPIC and ACT-R models.

Critics of the rational approach to understanding cognition argue, in line with Simon’s bounded rationality, that the optimisation of decision problems is too difficult and that it is unrealistic to expect human cognition to be able to achieve them (Gigerenzer & Todd, 1999). This critique perhaps misunderstands the goal of a rational explanation, which is to understand why the behaviour takes the form it does rather than the algorithm that
produces it. To use Marr’s (1982) example, rational analysis (Marr’s computational level) seeks to explain why a bird flaps its wings rather than how it flaps them. This response to the critique has been around for some time, Friedman and Savage (1948) give the example of hypothesising about an expert billiard player’s next shot by using complex formulas that account for ball positions, scoring and the physics involved.

It would in no way disprove or contradict the hypothesis, or weaken our confidence in it, if it should turn out that the billiard player had never studied any branch of mathematics and was utterly incapable of making the necessary calculations: unless he was capable in some way of reaching approximately the same result as that obtained from the formulas, he would not in fact be likely to be an expert billiard player (Friedman & Savage, 1948, p. 298)

In both rational analysis and computational rationality, the development of rational solutions to goals is a methodological tool that the experimenter or theorist uses to determine what external pressures may be acting on, and shaping cognitive mechanisms (Chater & Oaksford, 2000). In the case of rational analysis these pressures are presumed to result from the goals we aim to achieve and the environments in which they are enacted. In the case of computational rationality the external pressure comes from the attempt to maximise utility given the constraints imposed by the task, environment and the cognitive constraints. It is not the case that the participant has to resolve the optimisation problem as it is expressed when determining the optimal solution to the problem. The participant might explore multiple strategies before determining the combination of their resources that leads to the highest reward. That human cognition is adaptive (it exhibits flexibility in strategy selection) is not controversial. The rational aspect of the above methodologies concerns identifying what the likely external pressures are that cognition is adapting to.

Nonetheless, whatever the goal is that a person is trying to achieve, there has to be an algorithm that processes the information, and there has to be an underlying physical implementation of that algorithm in the brain. Marr (1982) argued that the different levels of explanation should be considered together and are mutually informative. The computational rationality approach to cognition achieves this by simultaneously considering both cognitive bounds and the overall goal of the person. An alternative but similar approach to this question of how to bridge Marr’s (1982) computational, algorithmic and implementation levels involves using rational process models (Sanborn, Griffiths, & Navarro, 2010; Griffiths, Vul, & Sanborn, 2012). Rational process models assume that where behaviour matches that of computational level probabilistic models, a good candidate for the algorithms that achieve this behaviour are the known best algorithms in computer science and statistics.

This aim of understanding the interaction of computational and algorithm constraints is shared with the computational rationality approach. The contrast perhaps lies in rational
process models starting from the computational level and then seeking to identify candidate algorithms, whilst computationally rational explanations focus on an assumption of cognitive constraints a priori, and then ask how they would combine under the external pressure of the computational goal.

Given the bounds imposed by a PhD, the thesis will primarily focus on applying a computational rationality analysis to a single illustrative phenomenon - the attraction effect. The effect highlights the unexpected impact that the changing composition of a choice set can have on which alternative is preferred. The attraction effect is another example of decision making where it is known that people do not behave in an optimal fashion. The effect has had a profound impact on the decision making literature since it so simply demonstrates a violation of rational value-maximising accounts. Since its publication in 1982 it has been cited over 1000 times, and over 50 times in 2014 (Google Scholar). Identifying the computationally rational approach to attraction effect choice sets reveals a more insightful understanding. This approach explains the effect in terms of why it is an advantageous thing to do rather than one that is simply irrational or the side-effect of bounded rationality.

The structure of the rest of this introductory chapter is as follows. It starts necessarily in Section 1.2 with a definition of the attraction effect, how robust it is and how it has been probed to date. Section 1.3.1 then outlines what the supposed rational behaviour should be in this context and why the attraction effect provides evidence that people do not do this. Section 1.3.2 outlines some of the major models that explain the effect with a view to identifying where the research in this thesis fills a gap. Finally, in Section 1.4 three research objectives are stated, centred on the extent that the attraction effect provides a challenge to rational theories of cognition. These objectives are addressed in the four content chapters that are papers currently under review, or about to be submitted.

### 1.2 The Attraction Effect

For those of us on a budget, trading off price and quality is a common dilemma. A high-quality product is appealing, but the price is often off-putting. A cheap alternative resolves the cost issue but you might worry about the quality. Huber et al. (1982) found that people in this situation can be biased toward favouring one or other of the alternatives, a phenomenon that has come to be known as the attraction effect. Given a choice where it is necessary to trade-off two attributes, the likelihood of choosing an alternative can be increased by adding an obviously inferior third option. For instance, people are more likely to choose an expensive high-quality alternative (A) over a cheap lower quality alternative (B) when a similar but more expensive and worse quality option than A is added to the choice set.

The attraction effect is not restricted to price and quality. Any two dimensions on
Figure 1.1: The attraction effect. In a choice between A and B described on two desirable attributes X and Y, the position of the decoy determines which alternative is the target and which is the competitor. If the decoy were in the dashed area Option B would be the target and Option A would be the competitor. The target option is expected to gain choice share relative to the competitor when a decoy is added to the choice set.
which alternatives might be described would, in theory, elicit the attraction effect. In their original demonstration of the effect, Huber et al. (1982) created choice sets that described restaurants in terms of distance and food quality, TV sets in terms of picture distortion and set reliability, as well as several other product categories described on various attributes.

The alternative that dominates the new inferior option is called the target, the added inferior option is called the decoy, and the remaining option is called the competitor (see Figure 1.1). The target and competitor are so-called because adding the decoy to a choice set is expected to increase the target’s choice share at the expense of the competitor. The attraction effect can be defined more formally in terms of ordinal relations. On one attribute the target is greater than the decoy which is in turn greater than the competitor. On the other attribute the competitor is greater than the target, which is in turn greater than the decoy. This means that overall, the target must be better than the decoy, but it is still ambiguous as to whether the competitor is better than either the target or the decoy. Formally, given two attributes \( X \) and \( Y \), and three alternatives \( t, c, d \) the ordinal relations must be such that:

\[
Y_t > Y_d > Y_c \\
X_c > X_t > X_d
\]

(1.1)

where \( X_t > X_d \) indicates that the value of the target option \( t \) on attribute \( X \) is greater than the value of the decoy option \( d \) on attribute \( X \).

Since the original demonstration in 1982, the attraction effect has been shown to be remarkably robust, and has been replicated in a number of different settings including: choice among political candidates (Pan, O’Curry, & Pitts, 1995), choice among gambles (Wedell, 1991), and choice among gambles with incentives (Herne, 1999). The attraction effect has been shown in honeybees and gray jays (Shafir, Waite, & Smith, 2002), and in slime mould (Latty & Beekman, 2011). The effect has also been replicated using within subjects designs (e.g., Herne, 1999; Trueblood, Brown, Heathcote, & Busemeyer, 2013; Soltani, De Martino, & Camerer, 2012).

Manipulations in studies on the attraction effect have included the position of the dominated option in the two-dimensional space (Huber et al., 1982; Wedell, 1991), requiring participants to justify their choices (Simonson, 1989), choosing among high and low quality brands (Heath & Chatterjee, 1995), familiarity with choices (Ratneshwar, Shocker, & Stewart, 1987), participant age (Kim & Hasher, 2005) and blood glucose levels (Masicampo & Baumeister, 2008).

While adding a decoy alternative that is strictly dominated by an existing option elicits the attraction effect, adding decoys with other dominance relations can elicit different effects. Three main contextual preference reversal effects have been identified. The attraction effect (Huber et al., 1982), the similarity effect (Tversky, 1972), and the compromise effect (Simonson, 1989). All of these effects relate to decisions between options which
involves trading-off two dimensions. In each case the proportion of choices between two options can be altered simply by adding a third option that also varies on the same two dimensions. The positioning of the third option is what determines the type of effect. As shown in Figure 1.2, in a choice between two alternatives A and B, the relative choice share of A over B can be increased by adding option m to elicit the compromise effect.

The addition of option m at the opposite ends of the attribute ranges to option B makes A a compromise between m and B. In the similarity effect, the new option s takes a disproportionate share away from B (the option it is similar to) so that A gains share relative to B.

The attraction effect is perhaps the most intriguing of the three effects because the added alternative is not a viable alternative by virtue of being strictly dominated by the target on both attributes. On the face of it, this obvious inferiority should make the decoy irrelevant since we expect it never to be chosen (other than perhaps through inattention). This is therefore a striking demonstration of context since the change in choice preferences is due solely to the presence of the decoy, which does not take up choice share itself (unlike the compromise and similarity effects). In this sense the attraction effect is the most challenging of the three preference reversals to explain, and therefore it is the focus of this thesis. The other context effects, however, are returned to in Chapter 4.

The attraction effect has been shown to be a robust phenomenon in multi-attribute decision making. The presence of an obviously poor alternative leads people to decide
more often in favour of the alternative that is clearly its superior. This finding has important theoretical implications for how we understand the cognitive psychology of decision making. The following section develops this theme showing how the effect violates rational models of decision making.

1.3 Theoretical Perspectives on the Attraction Effect

This section provides an overview of how the attraction effect is interpreted and explained in the decision making literature. The focus of this section is the common observation that rational models of decision making cannot explain the attraction effect. It is made clear why rational models cannot explain the effect, and then the focus turns to the consequences of interpreting the attraction effect in this light.

1.3.1 A Violation of Rational Axioms

Many authors (Ariely & Wallsten, 1995; Heath & Chatterjee, 1991; Huber et al., 1982; Louie, Khaw, & Glimcher, 2013; Rattaneshwar et al., 1987; Roe, Busemeyer, & Townsend, 2001; Sen, 1998; Simonson, 1989; Tsetsos, Usher, & Chater, 2010; Tversky & Simonson, 1993; Usher & McClelland, 2004) note that the attraction effect implies a violation of two axioms needed in rational models: Regularity and independence from irrelevant alternatives (IIA). This is important because it rules out a class of parsimonious, simple and rational explanatory models of human decision making known as value maximising models. Value maximising models stipulate that people should place a value on each alternative and then choose the alternative with the highest valuation. In Luce’s (1959) choice axiom this process is probabilistic such that the highest valued choice is the most likely to be chosen and each option’s probability of being chosen is a function of its valuation. This is a deceptively simple and intuitively plausible model of decision-making. In essence, you determine how much you value each alternative and select the one with the highest valuation. Given this apparent simplicity, and that this is a normatively rational model, it is all the more surprising that this is not what people do.

Independence from irrelevant alternatives is a necessary condition for value maximisation and stipulates that a decision maker must assign a value to each alternative based solely on that alternative’s own qualities. This means that an alternative will always receive the same valuation regardless of what other alternatives are also on offer. Formally, IIA means that:

\[
\frac{p(a|\{a,b\})}{p(b|\{a,b\})} = \frac{p(a|\{a,b,c\})}{p(b|\{a,b,c\})}
\]

where \(p(a|\{a,b,c\})\) is the probability of choosing alternative \(a\) given a choice set comprised of \(a, b\) and \(c\). Regularity requires that increasing the size of a choice set does not
increases the probability that any of its members are chosen. Formally:

\[ p(a|\{a, b\}) \geq p(a|\{a, b, c\}) \]

The attraction effect violates both of these axioms. Because the attraction effect means that the target is chosen more often when the decoy is added to the set we can see that it violates regularity. The probability of the target being selected from the smaller choice set is less than its probability of being chosen from the larger choice set: \( p(t|\{t, c\}) < p(t|\{t, c, d\}) \). The effect also means that the ratio of choices: target over competitor, is not the same when the decoy is absent versus when it is present, thus violating IIA:

\[ \frac{p(t|\{t, c\})}{p(c|\{t, c\})} < \frac{p(t|\{t, c, d\})}{p(c|\{t, c, d\})} \]

The attraction effect therefore, shows us that people do not value each alternative independently; a person’s preference for the target over the competitor will depend on whether the decoy is present or not. It is this contextual influence (the composition of the choice set) that means a person may preference reverse, preferring option \( a \) over option \( b \) in one context (decoy dominated by \( a \)), but \( b \) over \( a \) in another context (decoy absent, or decoy dominated by \( b \)).

In summary, exhibiting the attraction effect means that people do not, solely at least, use a value maximising decision making strategy. The observation that human, and other species’, decision making violates axioms of rational decision making models means that other models must be sought to explain how and why the effect might occur. The following section analyses some of the key models in the decision making literature that can reproduce the attraction effect. What is interesting about the majority of these explanations is that they implicitly focus on how the attraction effect might be produced rather than why it is produced. Although not stated explicitly, it seems that the violation of value maximisation implicit in the attraction effect has lead theorists to abandon the why explanation focussing instead on descriptive models that offer plausible mechanisms and algorithms.

### 1.3.2 Mechanistic Accounts of the Attraction Effect

This section focusses on the consequences of interpreting the attraction effect primarily as a violation of rational axioms. Models of the attraction effect, and other contextual effects, have predominantly (but not exclusively) focussed on identifying heuristics and or procedural decision making models that are capable of producing the effect. This inherently means that these models are descriptive rather than explanatory.

Several heuristics were among the first proposed explanations for the attraction effect.
One instance of these, discussed by Huber et al. (1982) was Russo and Dosher’s (1983) majority of confirming dimensions (MCD) heuristic. According to this explanation people compare each alternative on both dimensions and select the alternative with the most wins. Another explanation was provided by Simonson (1989) who argued that both the attraction effect and the compromise effect occur because they provide a reason for people to justify their choice. In the case of the attraction effect the dominance of the target option provides the reason. In the case of the compromise effect, the fact that, as the name suggests, the chosen option involves the least severe trade-off of the two attributes is the reason.

Huber et al. (1982) also considered the possibility of range and frequency effects, concepts first established by Parducci (1974) in relation to psychophysical judgements. According to this model, if the decoy option is positioned such that it is inferior to the target on the same attribute as the target is superior to the competitor, then the decoy has increased the frequency of alternatives that dominate the competitor on that attribute. Consequently participants might perceive an enhanced distance between competitor and target on that dimension. A range position is one in which the decoy is inferior to the target on the dimension on which the competitor is better than the target. This has the effect of extending the range of values on which the target option is weakest thereby making the loss on that dimension from competitor to target seem less extreme. A range-frequency positioning happens when decoys are inferior to the target on both dimensions simultaneously.

Both Huber et al. (1982) and Wedell (1991) have tested the impact of decoy positions on the size of the attraction effect. Huber et al. (1982) obtained a larger attraction effect when the decoy was positioned in the range extension position although this was not dependent on the amount of range extension. The Wedell (1991) study manipulated the position of the decoy within subjects and reported significant attraction effects for all positions of the decoy: 20.8% change for range, 22.7% change for frequency and 17.5% change for range-frequency. In an extra manipulation, Wedell tested a range decoy which was dominated by both the competitor and target options. According to range frequency explanations this decoy should still produce the attraction effect despite its being symmetrically dominated. This is because it still extends the range along the dimension on which the target is inferior to the competitor. Results showed no effect of this decoy, and hence Wedell (1991) argues that heuristic explanations such as having a justification (Simonson, 1989), or using MCD (Russo & Dosher, 1983) provide better explanatory power than the range frequency approach.

The above high-level heuristic accounts can reproduce the attraction effect. They are however, as noted recently by Soltani et al. (2012), not formally defined and therefore hard to contrast meaningfully in terms of predictions that they might make for different stimuli values. One of the first models of the attraction effect to be more formally defined
was the context dependent advantage (CDA) model (Tversky & Simonson, 1993).

The CDA model has two components that can affect the extent of the predicted attraction effect. First is the background context which essentially determines the weight that is given to each attribute. A decision maker would establish this weighting via experience of previous choices in which the same attributes were traded off. The background context may have an impact on choice proportions but is not strictly necessary to explain the attraction effect. The attraction effect can still be modelled with the assumption that the decision maker has no background frame of reference.

The second component, which drives the attraction effect, is called the local context. According to this account a relative advantage is computed between each pair of alternatives by comparing their attribute values with each other. When an alternative suffers a disadvantage (has a smaller attribute value), that loss is weighted more heavily than a corresponding gain would be. The consequence in an attraction effect choice set is that the target option has a larger sum of relative advantages than the competitor since it loses only once in the pairwise comparisons, while the competitor loses twice (to the target and the decoy on one of the attributes). This explanation assumes loss aversion (Tversky & Kahneman, 1991) as a given.

One feature of CDA is that it is static, that is to say it predicts a final outcome but not what might happen during the actual time course of a decision. In contrast, the leaky competing accumulators (LCA, Usher & McClelland, 2001; Usher & McClelland, 2004) and multialternative decision field theory (MDFT, Roe et al., 2001) models described in the following section are dynamic, and they depict the attraction effect emerging as a function of the deliberation process. These models are described as neurocomputational (Usher, Elhalal, & McClelland, 2008) and are suggested to represent neurally plausible ways in which deliberation, or evidence accumulation, may occur until a decision is reached.

Both MDFT and LCA can be interpreted as multi-layered connectionist networks. In MDFT the decision maker is assumed to stochastically switch attention between attributes. Whilst attending to an attribute the values of each alternative are compared and contrasted feeding into a valence measure for each alternative in the choice set. The valence is computed as the activation for that alternative minus the average activation of the other alternatives. As time passes and deliberation proceeds, the leaky integration of these valences leads to a preference state for each alternative. Here, the preference states compete with one another through inhibitory connections, the strength of which is distance dependent. If the decision has a time constraint then the option with the highest preference is chosen at the decision time. Otherwise preferences develop until one is sufficiently large to exceed a decision threshold.

In MDFT the attraction effect occurs because the decoy option develops a negative preference state. This happens because it repeatedly suffers when compared with the average of the target and competitor. The negative preference state of the decoy feeds
through a strong inhibitory link (because of its proximity) to the target, and through a weak inhibitory link (because of its distance) to the competitor. MDFT is a linear model and as such permits that negative preference states inhibiting neighbours actually result in an enhanced (the two negatives cancel) preference for the neighbouring option. As a result the decoy boosts the target more than the competitor. This process is akin to how edge enhancement effects are thought to be created in vision (Busemeyer & Johnson, 2004).

LCA also stochastically switches attention between attributes, with each option receiving activation according to the stimuli attribute values. LCA differs from MDFT in the next step, where it calculates the differences between each of the option’s activation levels. These differences are transformed into preferences via a loss averse value function. The preferences now compete as in MDFT. However, in LCA inhibition is not distance dependent and has a lower bound at zero. These properties mean that the competition in the final layer is not the cause of the attraction effect, as it is in MDFT. In LCA the attraction effect is explained through the loss averse value function that transforms the calculated differences into preferences. Because the competitor suffers two large losses, one to the target and one to the decoy, whereas the target only suffers one large loss, the target ends up with the smallest comparative losses in a model where losses outweigh gains. In LCA the mechanism that drives the attraction effect is essentially the same as that in Tversky and Simonson’s (1993) CDA model.

More recently, several new models have sought to improve and expand upon MDFT and LCA in a variety of ways whilst also accounting for the attraction effect. Soltani et al. (2012) developed a new ‘range normalization’ model to account for context dependent choice including the attraction effect. In CDA, LCA and MDFT there is an assumption that every alternative in a choice set is compared; consequently Soltani et al. (2012) argue that these models become increasingly computationally demanding as the size of the choice set and the number of attributes increases. According to the range normalization model the set of attribute values might be encoded in neural firing rates which have known physical limitations (cannot go negative and have an upper limit). It is however known that the dynamic range of neurons can be adjusted in order to represent stimuli distinguishably. The adjustment of this dynamic range (hence ‘range normalization’) is the proposed mechanism for dealing with different sized choice sets and the underlying mechanism behind context effects such as the attraction effect. The model is less computationally demanding since only the most extreme values in a choice set are necessary to compute a range.

Other recent models in this vein include the multi-attribute linear ballistic accumulator (MLBA) model (Trueblood, Brown, & Heathcote, 2014) and the associative accumulation model (AAM) (Bhatia, 2013), both of which can also account for the three contextual preference reversal effects. The AAM model assumes that the accessibility of an attribute
increases if: there are alternatives that have extreme values on that attribute, if there are many alternatives that have values on that attribute, or if there is a particularly salient alternative that has a value on that attribute. AAM models the attraction effect through the presence of two alternatives with high values on the same attribute (target and decoy) leading to increased accessibility of the target’s primary attribute (the attribute on which it dominates the competitor).

Accounts of the attraction effect provided by neurocomputational models suggest an interpretation in which the effect is almost an inevitable consequence of the neural mechanisms we employ when making decisions. Soltani et al. (2012) state that “context effects are a natural consequence of the biophysical limits of the neural processing in the brain” (p. 10) and Usher et al. (2008) argue that contextual preference reversals “demonstrate a limitation of rationality in choice preference” (p. 297). These models certainly do not present the attraction effect as an adaptive outcome, but rather describe it as a side effect of decision making mechanisms.

The dominant explanations of the attraction effect focus on how it might be produced, rather than why it is produced. This is symptomatic of the importance attributed to the effect violating IIA and regularity. Once it had been accepted that the behaviour is not explained by rational models, the search for mechanisms that might produce it commenced. The fundamental problem with this approach, is not the search for potential algorithms, but the abandonment of rationality as a constraining factor in that search.

If the axiomatic violations implicit in the attraction effect mean people are not value-maximising, the question should be what are they doing, as well as how might the attraction effect be produced? An explanation in which the attraction effect is a sensible thing to do because it results in better decisions than would otherwise be the case is an explanation of why the effect occurs. Such explanations have more explanatory power since they establish a causal link between mechanism and behaviour that is absent in descriptive mechanism models.

The preceding sections have outlined the attraction effect, how it is frequently interpreted in terms of its violation of the IIA and regularity axioms and argued that the current dominant explanations of the effect are in fact descriptive rather than explanatory. A value-maximising explanation of decision making is inherently explanatory in that it provides an answer as to why behaviour takes the form it does - because it is adaptive. Section 1.4 argues that violations of IIA and regularity implicit in the attraction effect do not mean that we have to abandon rational explanations of human decision making. We can still develop rational explanatory models if we take into account the bounded nature of our cognition when we ask what the optimal decision is. This will allow us to reinterpret the attraction effect as an adaptive behaviour and provide a causal rather than descriptive explanation for its occurrence.
1.3.3 Other Accounts of the Attraction Effect

Not all models and explanations of the attraction effect can be categorised as descriptive. Stewart, Chater, and Brown (2006); Stewart (2009) provide an alternative account of the attraction effect in Decision by Sampling (DbS). In DbS alternatives are assessed in a series of binary comparisons within attributes. Given that the target dominates the decoy on both attributes while the competitor only dominates the decoy on one attribute, the target is the alternative that benefits most from the procedure, and is therefore most likely to be chosen. DbS to some extent bridges the mechanistic approaches of MDFT and the normative approaches of utility maximising models (Stewart, 2009) since it offers an account of how a decision proceeds, as well as arguing that there are rational elements of cognition, such as memory’s adaptation to the environment. A key contribution in DbS is that many observed phenomena such as risk aversion occur because of the distribution of gains and losses that people experience in their environment. This explanation does not appear to extend to the attraction effect. The attraction effect in DbS seems better described as a consequence of the decision process rather than an outcome that is in any way adaptive.

In an approach that explicitly seeks a rational explanation, Shenoy and Yu (2013) argue that the attraction, compromise and similarity effects result from a process of inferring a ‘fair market value’. According to this explanation the alternatives offered are assumed to be representative of the marketplace. This allows the decision maker to infer what a reasonable offering in that market would be. For the attraction effect, the addition of the decoy lowers the indifference curve between the two attributes such that the target option sits above it, while the competitor remains on the indifference curve. Consequently the target is perceived to be a better value alternative in the presence of the decoy. Shenoy and Yu’s (2013) model focusses on interpreting contextual preference reversals in terms of rational behaviour. People exhibit the effect because it is beneficial to do so. The principle difference from the approach we take here is that they do not explicitly consider the cognitive bounds under which a person must operate as a foundational principle in determining what the optimal behaviour should be. Furthermore their model concerns inferring market conditions while the approach we have developed focusses on achieving a more accurate estimate of the true underlying expected value estimate in choices between gambles.

1.4 A Computationally Rational Analysis of the Attraction Effect

The above sections outlined the nature of the attraction effect and existing interpretations of it within the decision making literature. This section identifies the research opportunity
that will be explored in the remaining chapters. Three research objectives are addressed in the remaining chapters, and are returned to in the Discussion.

Section 1.1 (Approaches to Rationality) outlined the theoretical motivation behind applying computational rationality to understanding the attraction effect. In summary, this approach to understanding cognition considers the cognitive bounds that people are subject to, as well as the task and environment constraints. It is an analysis of what is optimal under these conditions that allows us to understand cognition as rational. This approach focuses on retaining the explanatory power of a rational analysis while acknowledging that human cognition is indeed bounded (Howes et al., 2009; Lewis et al., 2014). Recent successes of this approach include demonstrations that bounds on working memory imply that the gambler’s fallacy is in fact not irrational given working memory constraints (Hahn & Warren, 2009). This approach has also been used to demonstrate that people choose the optimal point to interleave tasks when the cognitive bounds on their task performance are taken into account (Farmer, Janssen, & Brumby, 2011; Janssen, Brumby, Dowell, Chater, & Howes, 2011).

The aim of this thesis is to challenge the unquestioning use of the attraction effect as evidence for the failings of rational decision models in explaining human behaviour. This overarching aim is broken down into three research objectives detailed in the next section.

1.4.1 Objective 1: Establish the Consequences of Violating Rational Axioms

Although the literature is quite clear that the attraction effect violates rational value-maximising models, little has been done to explore the limits and consequences of these violations. A prime example is that all attraction effect experiments involving gambles have used alternatives that either have the same expected value or an expected value determined to be subjectively equivalent (Huber et al., 1982; Wedell, 1991; Herne, 1999; Soltani et al., 2012; Trueblood, 2012). It is curious that the decision making held up as illustrative of the failure of value maximising models is elicited in choice sets where value-maximisation is already guaranteed (you cannot fail to maximise expected value in a choice set where the only viable alternatives have identical expected values).

The demonstration of irrationality implicit in the attraction effect relies on a qualitative rather than quantitative accordance with rational behaviour. That is to say the attraction effect is a logical violation of value maximisation but not a practical one. Both Jarvstad, Hahn, Warren, and Rushton (2014) and Hahn and Harris (2014) make this point with regard to many demonstrations of behaviour that logically imply irrationality.

On the one hand, this makes for simple and compelling demonstrations; on the other hand, however, it does not allow assessment of how costly such
violations might actually be to people going about their everyday lives. (Hahn & Harris, 2014, p. 51, original emphasis)

In some decision making paradigms the rationality of performance is assessed against the efficiency of the decisions, i.e., quantitatively - what percentage of the maximum return does the decision maker achieve? The rapid pointing paradigms in Trommershäuser, Maloney, and Landy (2008) use an efficiency metric, whereas performance in attraction effect experiments is measured against a qualitative accordance with the IIA and the regularity axioms.

The extent to which the attraction effect can be used to elicit a quantitative violation of rationality is explored in Chapter 2 where we attempted to elicit the attraction effect using a rapid pointing task - a paradigm in which people have been shown to have near optimal performance against a quantitative measure. A complementary approach is used in Chapter 3 where participants were offered choices between gambles in an attraction effect configuration, but where we systematically manipulated the difference in expected value between target and competitor alternatives. These chapters therefore explore the consequences of the attraction effect (as determined by a quantitative measure of performance) and under what conditions the effect can be expected to appear.

1.4.2 Objective 2: A Computationally Rational Model

While Objective 1 is about establishing the boundaries and consequences of violating IIA and regularity, it will not provide an answer to why the attraction effect exists in choices that do have similar expected values. To answer this question, we asked whether the attraction effect can be understood as rational under constraints. In choice sets where the axioms of rational models are violated, it is perhaps because people have uncertainty in their estimation of the value of each alternative. In such scenarios any additional information provided by the context will help resolve uncertainty.

The second objective is to determine whether there might be a causal explanation for the attraction effect. The cognitive constraint of noise in expected value calculation means that an optimal decision maker should make use of the contextual information present in the ordinal relations among the stimuli. An integration of this information with a standard expected value estimate will result in higher expected value decisions, and in the attraction effect.

The assumption of noise in cognitive processes is the simplest of cognitive constraints that could be considered. If this constraint is sufficient to provide a rational explanation of the attraction effect then it can be argued that human behaviour is actually very close to optimal and well adapted to these decision making scenarios. Chapters 3 and 4 outline a high level and more formal description of this approach respectively.
1.4.3 **Objective 3: Derive Novel Predictions**

It may always be possible to develop a rational model of some behaviour by considering new constraints in much the same way that adding extra parameters to a model will increase the range of behaviours that it can accommodate. That a rational explanation can be found does not intrinsically mean that it is the correct explanation for a behaviour. One way to provide extra support for an explanation is to predict new phenomena and test whether these occur empirically.

In Chapter 5 we test a novel prediction that we derived from our computationally rational model. The model predicts that in choices between prospects, the attraction effect should be reduced in the loss domain. To my knowledge the attraction effect has not been tested in choices between prospects that lose money, although it has been extensively shown in choices between prospects in the gain domain (Wedell, 1991; Herne, 1999; Soltani et al., 2012).

Whilst the empirical question is interesting in itself, the main objective is to provide further evidence in support of the claim that the attraction effect is an adaptive behaviour helping to maximise gain (or minimise loss) rather than it being a demonstration of our inability to apply value-maximising strategies. Whilst it is interesting to know more about the attraction effect, it is perhaps more valuable to provide evidence for the utility of using a computationally rational approach.

1.4.4 **Methods**

Each of the content chapters contains a detailed methods section. However, Appendix A provides details on aspects of the methodology that are common to all the chapters, and will help the reader understand precisely how the attraction effect was measured. Appendix B provides the reader with an overview of expected value, which plays a pivotal role in both the experimental design and the modelling reported in this thesis. Where the following chapters discuss ‘the attraction effect ordinal relations’, the reader may find it useful to refer back to Equation 1.1 on page 20.
Chapter 2

The Attraction Effect in a Rapid Pointing Task

Abstract

We tested whether people exhibit the attraction effect when choosing between motor prospects. Participants were required to make choices between prospects where the probability of success was determined by their own motor noise in a rapid pointing task. We also provided a mathematically equivalent set of prospects using a standard descriptive paradigm as a manipulation check. Results show that people do exhibit the attraction effect in choices among motor prospects. The effect was also present in our equivalent descriptive prospects. We discuss the implications for our understanding of the attraction effect, and of optimality in motor decision-making.

Keywords: Attraction effect; Motor prospects; Preference reversals

2.1 Introduction

When choosing between alternatives people often allow context to influence their decision. An example is the attraction effect (Huber et al., 1982) which occurs when a choice between two options is influenced by the addition of an inferior option which itself is not chosen. The attraction effect is of particular interest in decision making because it reveals a violation of the regularity and independence from irrelevant alternatives (IIA) axioms required by rational value maximising models such as Luce’s (1959).

To elicit the effect, choices are described on two attributes that trade-off, for instance cars that trade-off acceleration and fuel-efficiency. Consider a person torn between car A with good acceleration, but low fuel-efficiency, and car B with poor acceleration, but high fuel-efficiency. The attraction effect tells us that offering an additional third option which is similar to A but slightly worse on both attributes might bias our decision maker toward choosing car A over car B.
Figure 2.1: The Attraction Effect. If a decoy prospect is added to the dashed area, the safe prospect will gain choice share from the risky prospect. If the decoy were instead added to the solid area, then the risky prospect would be the target, and would gain choice share from the safe prospect - now the competitor.

The effect has been demonstrated in a variety of product categories (Huber et al., 1982), in perceptual decisions (Trueblood et al., 2013) and in prospects (Wedell, 1991; Herne, 1999). For prospects the two attributes that trade-off are probability and value. Figure 2.1 shows a configuration in which the attraction effect would be expected to occur. The safe prospect is called the target because it strictly dominates the decoy on both attributes, making it more likely to be chosen than the risky prospect, called the competitor. Note that the competitor only dominates the decoy on one attribute.

The attributes of alternatives in attraction effect experiments tend to be explicitly described, for example a participant might be offered a choice between ‘an 83% chance of $12’ and a ‘50% chance of $20’ (e.g., Wedell, 1991). By contrast, a burgeoning literature is examining how people make decisions when the probability of a prospect is implicit in a person’s own motor noise (Trommershäuser et al., 2008; Wu, Delgado, & Maloney, 2009; Jarvstad, Hahn, Rushton, & Warren, 2013). In this new paradigm people often make decisions which are near optimal and maximize gain. In such tasks participants are asked to rapidly point at a target on a display. Because a short time limit is imposed, people cannot be completely accurate. The measured variability in their motor movement can then be translated into a probability that they will hit the target. When presented with prospects in this form people have been shown to aim for a mean end-point that maximizes the reward they can achieve (Trommershäuser et al., 2008).
This optimality holds in a variety of manipulations such as adding a penalty zone (Trommershäuser, Maloney, & Landy, 2003a), multiple penalty zones and configurations (Trommershäuser, Maloney, & Landy, 2003b), adding noise (Trommershäuser, Gepshtein, Maloney, Landy, & Banks, 2005) and in assessing which of two configurations has higher expected gain (Trommershäuser, Landy, & Maloney, 2006). However, failure to maximise gain has been shown when the complexity of the payoff surface is increased through adding penalty zones that are asymmetrical (Wu, Trommershäuser, Maloney, & Landy, 2006).

The reported optimality of people’s decision making in rapid pointing tasks contrasts with the violations of regularity and IIA that are elicited by attraction effect choice sets, violations which supposedly prevent value-maximisation. If decision making is indeed optimal in pointing tasks then the paradigm might reveal a resilience to the influence of context shown in attraction effect experiments. However, the attraction effect is an extremely robust phenomenon, which as well as being shown across a wide range of products, has also been shown in lower level perceptual decisions (Trueblood et al., 2013) and across a variety of species, including birds, honeybees (Shafir et al., 2002) and even slime mould (Latty & Beekman, 2011). Given the robust nature of the effect it may also persist among motor prospects, despite motor paradigms usually eliciting near optimal performance. The experiment we report here sought to test for the attraction effect by presenting people with motor prospects in an attraction effect configuration.

### 2.2 Method

Our method of creating motor prospects that are mathematically equivalent to explicitly described prospects, is adapted from Wu et al. (2009); Wu, Delgado, and Maloney (2011).
Figure 2.3: Example stimulus in the descriptive condition. Participants indicated which prospect they would prefer to play.

Figure 2.4: Example stimulus in the motor paradigm. Participants were told to indicate which of the three motor prospects they would play. The values in green indicate the amount that would be won if the target were successfully hit.
2.2.1 Participants

Sixty-one (eight male) undergraduate participants with a mean age of 20 ($SD = 2$) were recruited from the University of Manchester. Participants received course credit for taking part in the experiment. The experiment was granted ethical approval, and participants gave informed consent. Participants attended for one session of approximately 30 minutes.

2.2.2 Apparatus

A 19 inch touch-screen at a resolution of 1280 by 1024 pixels was used throughout the experiment. In the choice phase participants responded by pressing 1, 2 or 3 on the numeric keypad of a standard Windows keyboard. The experiment was created in the Python programming language and run on a Microsoft Windows 7 PC.

2.2.3 Design

Within subjects, we tested whether a safe motor prospect would be chosen more often when in the target configuration than in the competitor configuration (see Figure 2.1). Participants chose between prospects in which an alternative’s probability of success corresponded to the width of a target. The value of the prospect was displayed above the target (see Figure 2.4). Our dependent variable was the selection rate of the safe prospect. As a manipulation check we also used a descriptive presentation mode, in which participants chose between prospects where the probability and value of success were explicitly stated (see Figure 2.3).

As in previous studies with rapid pointing tasks (review in Trommershäuser et al., 2008) a time limit meant that participants could not be sure of hitting the target. Our participants completed a training phase prior to the choice phase in which we recorded their hit points to recover a distribution of their accuracy around a target center. This distribution of participants’ hit points was then used to create individualized stimuli in the choice phase. This allowed us to control the probability of success for each participant by varying the width of the targets they were presented with.

Participants chose between a safe prospect (wide target) and a risky prospect (narrow target). The safe prospect was set at .7 probability of winning £20. The risky prospect had a probability of .3 and a value of £75. A pilot study was conducted to determine a value that would make the risky prospect subjectively equivalent to the safe prospect. Decoys were always .15 less likely to win and of £5 less value than their dominating prospect. The attraction effect predicts that the safe prospect will be chosen more often when in a target configuration than a competitor configuration. Repeated presentation of the choice sets allowed us to determine a safe prospect selection rate in each configuration. In total there
were 27 motor trials and 27 descriptive trials. The 54 trials were presented in random order. The order of each alternative on the screen was also randomized.

2.2.4 Procedure

Training Phase

In the training phase participants learned their own motor noise in a rapid pointing task. Participants were instructed to touch a green bar on the left hand side of the screen, after which they had 500 msec to touch a yellow target bar on the right hand side of the screen (see Figure 2.2). The target zone was 20 pixels in width and 1025 pixels to right of the of the start bar. The start bar was 50 pixels wide and covered the full height of the display (1024 pixels), as did the target zone.

Participants completed 100 training trials as described above. Participants were not informed that they would be making decisions between different target widths in the subsequent choice phase. If participants successfully hit the target within the time limit the word ‘Hit’ would appear in green. If they were within the time limit but missed, the word ‘Miss’ was displayed in orange. If they exceeded the time limit, the message ‘Too slow’ was displayed in red and a warning sound was played.

Choice Phase

In the choice phase participants saw three prospects presented on the screen and a message stating ‘Evaluate’ above them (see figure 2.4). Participants were instructed to press the space bar when they were ready to indicate their choice. The message would then change to ‘Choose’, and one of three prospects would disappear. Participants had to choose between the remaining two prospects by pressing the appropriate number on the keyboard. In the majority of trials the decoy was removed when the space bar was pressed. However, in some trials the decoy remained and either the target or competitor prospects were removed. These trials were excluded from the analysis but were included in the stimuli to encourage participants to evaluate all the prospects. Participants indicated their preferred prospect using the numeric keypad, 1 for the left most prospect in the display, 2 for the middle prospect and 3 for the right-most prospect.

Participants only had two seconds to make their response after pressing the space bar. If they exceeded this time limit, ‘Too slow’ was displayed and the trial was discarded from the analysis. The process of evaluating, removing an option, and choosing under a time constraint was adapted from Soltani et al. (2012) with the purpose of forcing participants to take all three alternatives into consideration when making their choice.

Rather than present identical choices nine times, the values and probabilities were jittered such that values were either £19, £20 or, £21 and probabilities were either .69,
.70 or .71. This resulted in nine safe prospects with a mean value of .7 (£20). The same procedure was applied to the risky prospect and the decoy prospects.

In the choice phase participants were asked to indicate which prospect they would prefer, but they did not go on to play the prospect, nor receive any other type of feedback.

### 2.3 Results

Analysis of the training phase showed that the distribution of participants’ hit points was well described by a normal distribution centred on the target (see Figure 2.5). The mean hit point (across participants) was 1092 pixels (eight pixels to the left of the center of the target). The mean standard deviation was 27 pixels.

Figure 2.6 shows the rates with which participants chose the safe and risky prospects in each presentation mode and according to whether the safe prospect was the target or competitor. In the motor condition participants chose the safe prospect in 73% of trials when it was the target, and in 62% of trials when it was the competitor. In the descriptive condition the safe prospect was chosen in 70% of trials when it was the target, and in 57% of trials when it was the competitor.

Although we sought to make the risky prospect equally attractive, the results show that in fact the safe prospect was preferred in 63% of trials when there was no decoy present.

Table 2.1 shows the difference between the rate the safe prospect was chosen when it was the target minus the rate it was chosen when it was the competitor. The neutral row
Table 2.1: Proportion of trials where the safe prospect was chosen.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Motor</th>
<th>Descriptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>.73</td>
<td>.70</td>
</tr>
<tr>
<td>Neutral</td>
<td>.65</td>
<td>.63</td>
</tr>
<tr>
<td>Competitor</td>
<td>.62</td>
<td>.57</td>
</tr>
<tr>
<td>Preference Reversal Rate</td>
<td>.11</td>
<td>.13</td>
</tr>
</tbody>
</table>

Figure 2.6: Proportion of trials in which the safe prospect was chosen. Note that the safe prospect is chosen more often when it is the target - as predicted by the attraction effect. Error bars are standard error.

in table 2.1 indicates the rate that the safe prospect was chosen when there was no decoy present. In the motor condition the safe prospect was chosen significantly more often when it was the target than when it was the competitor, $t(60) = 4.36, p < .001, d = 0.34$. The effect was also significant in the descriptive condition $t(60) = 5.13, p < .001, d = 0.38$.

The above analysis was conducted on the rate that the safe prospect was chosen in each configuration. However, the analysis could have been carried out on the complementary choices of the risky prospect. In order to check that the results were not dependent on which prospect we analysed, we conducted the same analysis on the risky prospect and found the same result, $t(60) = 4.408, p < .001$. As in Wedell (1991), an arcsine transformation of the choice proportions was applied for statistical analyses.
2.4 Discussion

In the motor paradigm, participants chose a prospect more often when it dominated a decoy than when it did not. This difference was significant, indicating that participants did exhibit the attraction effect. This finding adds to the body of literature showing that the attraction effect is pervasive across tasks and organisms.

An objection to our design might be that participants did not have to make a pointing movement in the choice phase. They were simply asked which alternative they would select if it were offered. Consequently it could be argued that the training phase might have had no impact on the participants’ subsequent decision making. It is possible that we would have elicited the same results without motor training because participants would still have had to trade off the attributes of target width and outcome magnitude. However, our design is adapted from, Wu et al. (2011) who tested for differences in neural activity between a width judgement task and a hypothetical motor prospect task, using identical stimuli. They found that activity in the medial prefrontal cortex correlated with probability in the hypothetical motor decision task, but not with stimuli width in the size judgement task. This suggests that, given hypothetical choices between motor prospects, people can and do represent target width in terms of probability.

It is particularly interesting that we found the attraction effect using a motor paradigm because this paradigm has been shown to facilitate optimal value-maximizing decisions. In stark contrast, the attraction effect is usually cited as evidence that human decision-making does not meet normative standards of rationality. However, recent work shows that the apparent superior performance of people in motor tasks relative to classical decision tasks, might actually be due to how performance is measured differently in the two domains (Jarvstad et al., 2013). Performance in motor tasks is typically measured as a percentage of the total possible reward that is actually achieved (efficiency). In many classical decision tasks, optimal performance is determined as conforming to normative axioms. In a study using both motor and classical paradigms, Jarvstad et al. (2013) found no difference in performance when efficiency was the metric applied to both. Similarly, our own work has recently shown that the attraction effect can be attenuated by offering choices between prospects that differ in expected value, suggesting that when value-maximization is possible, people do so (Farmer et al., under review).

In the experiment reported here, we have used a motor paradigm but measured performance by conformity with normative axioms (IIA and regularity), and found a violation. Wu et al. (2009) also found that a motor decision task did not attenuate the violation of the independence axiom required in expected utility theory (Von Neumann & Morgenstern, 1944). Together, these findings suggest that the different performance (rational and non-rational) in different decision domains might simply reflect the way in which performance is measured. Therefore it is plausible that human decision making does not
vary in rationality from one paradigm to another, but that different measures elicit different findings. For instance, the efficiency measure typically used in pointing paradigms is unlikely to detect the attraction effect, precisely because the attraction effect occurs in choices between alternatives that have similar expected values.

However, this insight still does not explain why the attraction effect occurs. There is still an intriguing contextual influence despite the limited implications for efficiency. Interpreting the attraction effect as a violation of normative axioms reveals little about why people make use of contextual information. An alternative approach is to consider the constraints imposed by the task and human cognition, before asking what the optimal course of action is (Howes et al., 2009). Using this approach it is possible to show that exhibiting the attraction effect is an optimal behaviour when alternatives have similar utilities (Howes et al. submitted, see also Shenoy & Yu, 2013). As an example, consider a risk-neutral choice between prospects. The normative standard is simply to calculate the exact expected value of each prospect and choose the highest. However, people, unable to calculate expected value with absolute precision, also make use of the contextual information. Namely that the target option dominates on both attributes while the competitor dominates on just one. Thus in decisions where alternatives are valued similarly, but with some degree of uncertainty, it is beneficial to make additional use of the information provided by the context of the choice set.

In our model of the attraction effect contextual information is most useful when the utilities of alternatives are very similar. If instead, the choice involves options that have very different utilities then context will have little impact. This understanding of the attraction effect together with the findings of Jarvstad et al. (2013) and Wu et al. (2009) allow us to piece together a framework in which the effect is both optimal and domain independent. As Jarvstad et al. (2013) note, in tasks where optimal performance is measured as efficiency the alternatives must vary in expected value otherwise efficiency cannot be measured. In tasks where performance is measured as conforming to an axiom, decisions are typically made between alternatives that are hard to choose between (in the sense that they are roughly equally attractive), for example, the common consequence effect in Kahneman and Tversky (1979), violates the independence axiom, however the choices on offer differ in expected value by less than one percent. Similarly, attraction effect studies have used choices between prospects with identical expected values (Huber et al., 1982; Herne, 1999; Wedell, 1991). It is precisely because the attraction effect is elicited in choices between alternatives that are equally valuable that context becomes a useful source of information. This prediction is not dependent on whether the task is framed as a motor or classical decision task, and hence, we believe, why we have found the effect in decisions among motor prospects.
Chapter 3

The Attraction Effect as an Adaptive Response to Uncertainty

Abstract

The attraction effect (Huber et al., 1982) shows that adding a third decoy alternative to a choice set can alter preference between the original two members. This simple demonstration of context dependence is taken as evidence against a class of parsimonious value-maximising models, since in order to maximise value, such models must evaluate alternatives independently of one another. We reproduce the attraction effect using prospects and show that, with the assumption of noise in expected value calculation, the decoy provides information about the expected value of the other alternatives, such that selecting the alternative that dominates the decoy achieves a higher expected value. We establish a model that reveals the relationship between how easily a person can distinguish the expected value of two prospects and the extent that they should exhibit the attraction effect. In a series of experiments we show that participants exhibit the attraction effect less as the expected values of target and competitor prospects become easier to tell apart. We reproduce this finding using several paradigms, including low-level perceptual decisions and higher level decisions from description. These findings are predicted by our model and suggest that, far from preventing value-maximisation, people exhibit the attraction effect, precisely in order to maximise value. Keywords: attraction effect, expected value, preference reversal, value maximisation, bounded optimality

3.1 Introduction

The attraction effect (Huber et al., 1982) refers to the puzzling change in preference that occurs when an apparently irrelevant alternative is added to a choice set. This decoy option, despite not being chosen, nonetheless changes preferences between existing members of the set. Over thirty years of research into the effect have shown it to be a robust
phenomenon highlighting the impact of context on decision-making. It has been found in choices among prospects (Herne, 1999; Soltani et al., 2012; Wedell, 1991), in low-level perceptual decisions (Choplin & Hummel, 2005; Trueblood et al., 2013), in inference tasks (Trueblood, 2012), choices between consumer products (Huber et al., 1982; Simonson & Tversky, 1992), in animals’ choices (Shafir et al., 2002) and even in decisions made by slime mould (Latty & Beekman, 2011).

Perhaps the biggest influence of the effect on the decision making literature has been the observation that it violates the regularity and, or, independence from irrelevant alternatives (IIA) axioms necessary in Luce’s (1959) choice axiom and other simple scalable models (Ariely & Wallsten, 1995; Heath & Chatterjee, 1991; Huber et al., 1982; Louie et al., 2013; Ratneshwar et al., 1987; Roe et al., 2001; Sen, 1998; Simonson, 1989; Tsetsos et al., 2010; Tversky & Simonson, 1993; Usher & McClelland, 2004). This failure of rational models to account for the attraction effect often results in comparisons between human decision making and the supposed optimal behaviour: “people err by complicating rather than simplifying the task” (Tversky & Simonson, 1993, p.1188); “A limitation of rationality in choice preference” (Usher et al., 2008, p.297).

In this paper we test whether the attraction effect can in fact be interpreted as an adaptive behaviour given only the assumption that a decision maker has noisy computational processes. This assumption is sufficient to show that existing normative models of decision making are not well matched to the task, and in fact, that an agent exhibiting the attraction effect is likely to outperform such models. We develop a simple high-level model that chooses between prospects by supplementing an expected value calculation with a second estimate based on the ordinal relations among the alternatives on offer.

We then demonstrate that this model captures human performance in both high level decisions from description and lower level perceptual decisions. This model allows us to predict the size of the attraction effect using just the difference in expected value between alternatives and an agent’s accuracy in perceiving that difference. In what follows, we outline the attraction effect and its violation of the axioms of value-maximising models. We then present our model and the data from two experiments designed to test it.

3.1.1 The Attraction Effect and Value-Maximisation

In the standard form of the attraction effect, options in a choice set are described by two attributes that trade-off, such as the probability and value of prospects, or the fuel-economy and acceleration of a car. Two of the options are difficult to choose between because each dominates the other on one of the attributes. When a third decoy option is introduced, it is dominated by one of the original options (the target) on both attributes and by the other option (the competitor) on only one attribute, see Figure 3.1.

Adding this asymmetrically dominated decoy has the effect of biasing choice away
from the competitor toward the target. In a two-option choice, each option might be
chosen 50% of the time, but when the decoy is added, the target might be chosen 60% of
the time, and the competitor 40% of the time. Note that the decoy is not chosen as it is
obviously worse than the target. Despite this, its presence influences the ratio of choices
between target and competitor.

The primary reason this behaviour is of interest is that it represents a preference re-
versal. The same alternative can be more or less preferred, depending not on its own
properties, but on the context provided by the composition of the choice set. This in
turn is interesting because simple models of decision making such as Luce’s choice ax-
iom only allow us to be sure of choosing the best alternative when each item is valued
independently of its context. In Luce (1959) IIA requires that:

\[
p(t | \{t, c\}) = \frac{p(t | \{t, c, d\})}{p(c | \{t, c\})} = \frac{p(t | \{t, c, d\})}{p(c | \{t, c, d\})}
\]

where \( p(c | \{t, c, d\}) \) is the probability of choosing the competitor given a choice set com-
prised of target, competitor and decoy.

Because the attraction effect means that the target is chosen more often when the
decoyn is added to the set we can see that the ratio of choices: target over competitor, is
not the same when the decoy is absent versus when it is present, thus violating IIA:

\[
p(t | \{t, c\}) < \frac{p(t | \{t, c, d\})}{p(c | \{t, c, d\})}
\]

Consider a decision between three prospects such as Figure 3.1, where attribute 1
is outcome value, and attribute 2 represents percentage probability. An expected value
maximiser, consistent with IIA, would multiply the probability of each prospect by the
value. They would then simply choose the alternative that had the highest expected value.
This method is guaranteed to achieve the highest possible expected value. If instead we
applied a heuristic consistent with the attraction effect, whereby the chosen alternative is
the one that accumulates the most wins along each attribute(target = 3, competitor = 2,
decoyn = 1), then the best alternative might be chosen, but that could not be guaranteed.
There are many scenarios consistent with the attraction effect in which the competitor
could have a higher expected value than the target.

A related principle in many rational models of decision making is regularity. This
states that increasing the set size cannot increase the probability of any of its original
members being chosen. The attraction effect is a clear violation since the probability of
the target option being selected increases when the set is expanded by adding a decoy
alternative: \( p(t | \{t, c\}) < p(t | \{t, c, d\}) \).

Previous demonstrations of the attraction effect, and of the above axiom violations,
use decisions between alternatives that are approximately equally valuable. Indeed, many
Figure 3.1: The attraction effect. The decoy is dominated by the target on both attributes, but by the competitor on only one. In this figure option A is the target and option B is the competitor. If the decoy were in the solid area rather than the dashed area, then option B would be the target and A the competitor.
designs explicitly use target and competitor prospects that have the same expected value (Huber et al., 1982; Wedell, 1991; Herne, 1999), or that have subjectively equivalent expected values (Soltani et al., 2012). The attraction effect has also been elicited in size judgements (Trueblood et al., 2013) between rectangles that had the same area.

This feature of these designs is intended to provide the most favourable environment for eliciting the attraction effect. Having the same expected value makes the options hard to choose between, so preferences may be less certain and more easily biased by the decoy. Indeed there is some evidence for this as Mishra, Umesh, and Stem Jr. (1993) show that initial preference levels can predict the extent of the attraction effect.

It is a curious aspect of the existing literature on the attraction effect that it is typically demonstrated in choices between alternatives that are equally valuable. Given that the effect is interesting because it violates axioms that underpin value-maximising models, it seems important to understand whether the effect persists when a person is presented with alternatives that are not equally valuable, and therefore making the wrong choice has some consequence for the decision maker. We investigate this question empirically using choices between prospects that vary in expected value.

If the attraction effect does not persist when people can perceive the difference in expected value between two prospects, then the violation of axioms in value-maximising models only occurs when value-maximisation is already guaranteed, because the viable alternatives have the same value. Why though, should the attraction effect occur when alternatives are difficult to tell apart? To answer this question we present a high-level model that suggests the attraction effect is actually an adaptive response to this uncertainty in determining the best alternative. We argue that the exhibiting the attraction effect results in choices that achieve a higher expected value than would be the case if people did not exhibit the effect and their choices were consistent.

3.2 A Model of the Attraction Effect as an Adaptive Response to Uncertainty in Expected Value Calculation

In this section we develop a simple high-level model, with the goal of illustrating that the attraction effect is an adaptive response to inherent uncertainty. This approach stands in contrast to some existing models that suggest that contextual preference reversals might emerge as a result of how the decision-making system is implemented neurally (Roe et al., 2001; Soltani et al., 2012; Usher & McClelland, 2004). Such models provide convincing accounts of how the attraction might effect occur. Our approach, instead seeks to show why the effect is adaptive. The model we present here shows that by calculating expected value and updating that estimate with information provided by the ordinal relations in an attraction effect choice set, a decision maker will both exhibit the attraction effect and
achieve higher expected value than would otherwise be the case.

Our founding assumption is simply that people cannot perform perfectly accurate expected value calculations. When presented with a prospect in the form of ‘A 17% chance of 59 points’ we assume instead that a person’s belief regarding the expected value of the prospect is described by a normal distribution centred on the true expected value. This reflects the fact that a person may be uncertain about their expected value calculation allowing for the possibility that they may have over or underestimated.

When presented with two prospects to choose between, a person would have two such distributions, one for each alternative, as in the left panel of Figure 3.2. The extent to which these distributions overlap would then determine the probability with which we expect that person to choose either alternative. This can be represented by the distribution of differences in expected value, as in the right panel of Figure 3.2, where the area to the right of the dashed vertical line represents the probability that prospect A will be perceived to have higher expected value than prospect B.

If the mean of the distribution of differences is near zero, there is there some uncertainty as to which alternative is best. This uncertainty peaks for any pair of alternatives that are perceived to have identical expected values, i.e., the difference distribution is centred on zero. Alternatively, a person may be certain which alternative is best if the distribution of the differences only encompasses zero in the extreme tails. The above process is simply a probabilistic expected value maximisation. However, in choices between prospects that are described on two attributes (probability of success, and value of success) there is additional information to be gleaned from the ordinal relations along the probability and value dimensions as demonstrated in the following section.

3.2.1 Estimating Expected Value With Ordinal Relations

The attraction effect is defined by ordinal relations such that the target must strictly dominate the decoy on both attributes, and the competitor on one attribute. The competitor must dominate both the target and the decoy on one attribute only (see Figure 3.1). These ordinal relations actually provide more information than a simple ‘target expected value must be greater than decoy expected value’ heuristic.

If the ordinal relations were the only information available it would still be possible to make a choice. This is because the strict dominance of the target over the decoy limits the possible expected value orderings of the three alternatives. Consider the following example in which option A is a target prospect that dominates the competitor on the probability attribute. The attraction effect stipulates that: $p(A) > p(D) > p(B)$ and that: $v(B) > v(A) > v(D)$, where A is the target, B the competitor and D the decoy. The expected value of these prospects is given by the product of their attributes, and because option A dominates D on both attributes, the expected value of A must be greater than
Figure 3.2: Noisy perception of expected value. The left panel represents the uncertain perceived expected values of two prospects, A and B. The right panel represents the perceived uncertain difference in expected value between A and B. The probability of choosing prospect A is determined by the area to the right of the vertical dashed line. The area to the left of the line is the complementary probability that B will be chosen.

that of D. This leaves just three possible expected value orderings, in two of which A has a higher expected value than B:

$$E(A) > E(D) > E(B)$$

$$E(A) > E(B) > E(D)$$

$$E(B) > E(A) > E(D)$$

Consequently, the probability that the expected value of A is greater than B is two thirds, and there is certainty that the target expected value is greater than that of the decoy. Note that in addition to inferring that the target must be better than the decoy, we have now shown that it is also possible to infer the probability that the target is better than the competitor.

If the distributions of probability and value from which the prospects were drawn are known, it is possible to measure the increase in expected value relative to the competitor that results from being a target alternative. By taking random samples from the probability and value distributions, then analysing the expected value of those prospects that conform to the attraction effect ordinal relations, it is possible to estimate the average expected value of targets, competitors and decoys. Figure 3.3 shows the distribution of differences between target and competitor when sampled according to the attraction effect ordinal relations. Note that the mean of the distribution is greater than zero, meaning that, on average, the target has a higher expected value than the competitor.
Figure 3.3: The distribution of differences in expected value between target and competitor prospects for a given environment. In this figure, three sets of 1m normally distributed values ($\mu = 20, \sigma = 5$) were randomly paired with three sets of 1m probabilities (Beta distributed, ($\alpha = 2, \beta = 2$). The triplets of prospects that conformed to the attraction effect ordinal relations were then used to plot the difference in expected value between the target and competitor.

This analysis shows that the ordinal relations of the three attraction effect alternatives on two dimensions can be used to estimate the average expected value of target, competitor and decoy options in that environment. Crucially, the attraction effect ordinal relations imply that the target alternative is likely to have the highest expected value. In the following section we examine the consequences of combining a noisy expected value calculation with the additional information provided by the ordinal relations.

### 3.2.2 Predicting the Attraction Effect

We contend that the attraction effect occurs when there is uncertainty as to whether the target or competitor prospect has higher expected value. This uncertainty is reduced by taking into account the likely differences in expected value implied by the ordinal relations. The result is simply that the perceived difference distribution (target - competitor) is shifted in favour of the target. The consequence of this is that the same alternative will be chosen more often when it is in a target configuration than in a competitor configuration, or when there is no decoy present.

Figure 3.4 schematically shows how the attraction effect occurs according to the model. In each panel the dashed line represents the perceived difference between two

---

1This analysis is robust, it does not matter whether the attribute parent distributions are normal, uniform, beta or a combination thereof, the same qualitative effect is found. In the example in Figure 3.3, the size of the effect is in fact dependent on the coefficient of variance in the value distribution. Larger variance relative to the mean results in a greater expected value for the target relative to the competitor.
prospects (A minus B). The mean of this distribution is centred on the true difference between the prospects. The solid line shows the perceived difference once a person has taken into account the extra information provided by the ordinal relations. Since the ordinal relations imply that the mean of the difference distribution (target - competitor) should be greater than zero (see Figure 3.3), the updated perceived distributions have simply had their mean shifted by one standard deviation toward the target alternative.

The left column shows the effect when A is the target and the right column when B is the target. The shaded area represents the proportion of choices for prospect A which is larger when A is the target, the non-shaded area shows the proportion of B choices which is larger when B is the target.

According to the model, difference in expected value between target and competitor prospects should moderate the size of the attraction effect. If there is a large difference in expected value it is more likely that the difference distribution will only encompass zero in the tails. Therefore a shift in the perceived difference in favour of the target will result in a smaller change in choice proportions compared to a situation where the two alternatives have similar expected values. This is illustrated by the smaller change in the shaded area that occurs in row two of Figure 3.4 compared to row one. In other words, if you already perceive an alternative to be much better, additional information to that effect won't make much difference. If on the other hand you find two alternatives very difficult to choose between, additional information can have a large effect.

The second, related prediction, is that increased certainty in expected value calculation will mean that a smaller difference in expected value will be necessary to attenuate the effect. For a given difference in expected value, if the variance on the difference distribution is low it is even more likely that zero will be in the tails. This is shown in the third row of Figure 3.4, the lower variance means that the change in the shaded area is less than in row two despite both rows having the same expected value difference. Put simply, the more accurate you are at perceiving differences, the more limited the range of scenarios in which additional information is useful.

To summarise, this model predicts that the size of the attraction effect is determined by the difference in the mean expected value of target and competitor, and by the accuracy with which people can perceive the expected value of the target and competitor. This is essentially an effect size calculation on the hypothesis that target and competitor have different utilities. The smaller this effect size the larger the attraction effect. On the other hand, if there is a large effect size, and a person perceives that the target and competitor

---

2The means of the difference distributions have been shifted by one standard deviation to reflect the fact that the more reliable an expected value difference estimate is, the less useful additional information can be. If a decision maker is nearly certain of the difference between two alternatives then they would update their difference distribution by a proportionally small amount. One standard deviation is used in Figure 3.4 in order to provide a schematic illustration of the effect, in practice a shift by a smaller proportion of the standard deviation produces the effect size observed in the attraction effect literature.
Figure 3.4: The attraction effect and expected value difference. The dashed line shows the perceived difference distribution of prospect A minus prospect B. The solid lines show the difference distribution with a mean shifted by one standard deviation toward the target (A in the left-hand column, and B in right-hand column). In each panel, the shaded region shows the proportion of choices for prospect A. Along each row, the difference between the shaded areas represents the size of the attraction effect. In the second row the perceived difference in expected value is greater than that in the first row, resulting in a smaller attraction effect. In the third row the variance in expected value calculation has been reduced, also resulting in a smaller attraction effect relative to row two, despite both rows having the same difference in expected value.
are very different there will be no attraction effect. In the following experiments we test the model, Experiment 1 tests the effect of absolute expected value difference on choices between alternative prospects. In Experiment 2 we extend this test to further paradigms and use a relative difference in expected value.

3.3 Experiment 1

3.3.1 Methods

Participants

Forty-seven students and staff (14 male) from the University of Manchester volunteered to take part. Participants were between 21 and 53 years old \((M = 27)\). Informed consent was collected and participants were paid £7.00. The experiment took approximately 45 minutes to complete.

Design

We used a within-subjects design to test the effect of expected value difference on the dependent variable of preference reversal rate. For each participant preference reversal was measured as the rate that an alternative was chosen when it was the target minus the rate that it was chosen when it was the competitor. This measure is based on that used by Wedell (1991). Note that a rate could be calculated since each unique choice set was presented eight times.

The independent variable of expected value difference had four levels (zero, three, six and nine). In order to achieve four levels of expected value difference, eight prospects were chosen. Four (called the V prospects) were given a fixed probability of 0.2 and varying values to achieve the expected values 8, 11, 14 and 17. The other four prospects (the P prospects) were given a fixed value of 25 and varying probabilities to achieve the same expected values as the V prospects, see Table 3.1. Each V prospect had a higher value, but lower probability than each P prospect. Each of the V prospects was paired with each of the P prospects to create 16 choice sets. Prospect pairs with the same expected value had an expected value difference of zero, while, for instance, a prospect with expected value eight paired with a prospect with expected value 17 had an expected value difference of nine.

Each of the choice sets was presented once with prospect P as the target and once with prospect V as the target resulting in 32 unique choice sets. Each of these unique choice sets was repeated eight times creating a total of 256 trials per participant. The experiment was divided into four blocks of 64 trials. The presentation of the trials was completely randomised with the exception that two trials could not be consecutive if the
only difference between them was the position of the decoy. This control was added to ensure that the experimental manipulation was not too obvious to the participants. The decoy position defined whether prospect V or prospect P was the target. The decoys were always of the same value as their target but with 10% lower probability.

**Stimuli**

Participants were asked to choose between three prospects. Each prospect had a probability $p$ of winning a value $v$ in the form $p(v)$ or $(1-p)(0)$. The probability of each prospect was presented to participants using a grid consisting of 100 squares (see Figure 3.5). For a prospect with a success probability of 0.6, 60 of the squares were shaded green. For a success probability of 0.4, 40 of the squares were shaded green, and so on. The position of the green squares within the grid was randomised. The value of each prospect was presented in an identical 10 by 10 grid immediately below the probability information. For the value grid, the display was not randomised, and was shaded red. Forty red squares indicated a value of 40, 70 red squares a value of 70, and so on. There were six grids in total, one probability and one value grid for each of the three prospects.

**Procedure**

At the start of each trial all of the displays were blank. Participants revealed the probability and value information for each prospect by holding the mouse button down in a blue shaded area between the probability and value displays. Lifting off the mouse button cleared the display. Participants could only view one prospect at a time, but could choose any order to view them in. This prevented participants from assessing alternative prospects by making a perceptual comparison of the visual density of the different displays, encouraging them instead to encode the information. In each trial, participants were given five seconds of viewing time that they could distribute between each of the prospects as they saw fit. A five-second countdown timer was displayed on the left of the interface. The timer only counted down while the participants had the mouse button depressed, and the probability and value information were visible. Participants could only

---

Table 3.1: Expected value difference between prospect pairs.

<table>
<thead>
<tr>
<th>High probability prospects (P)</th>
<th>Low probability prospects (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.2(40)</td>
</tr>
<tr>
<td>.68(25)</td>
<td>9</td>
</tr>
<tr>
<td>.56(25)</td>
<td>6</td>
</tr>
<tr>
<td>.44(25)</td>
<td>3</td>
</tr>
<tr>
<td>.32(25)</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 3.5: The stimuli used in Experiment 1. The density of green squares (top row) represents the probability of the prospect. The number of red squares (bottom row) represents the value of the prospect. Each column represents an alternative prospect. The probability and value of only a single prospect are displayed at any one time. The participant chooses a prospect by pressing the corresponding 'select' button.

make their choice once the timer had reached zero. An enforced assessment duration allowed us to control the amount of effort participants put into each trial. Participants were asked to choose the prospect they preferred by clicking a button marked ‘Select’ that was positioned below the value grid for each of the prospects. Participants did not receive any feedback on their decisions.

3.3.2 Analysis

As noted, in order to determine the preference reversal rate, we subtracted the rate that P prospects were chosen when they were the target, from the rate that they were chosen when they were the competitor. This metric was averaged across the prospect pairs for each level of expected value difference. A positive difference indicated that participants preferred a prospect more often when it was the target than when it was the competitor; this is the outcome that is expected when measuring the attraction effect. A negative difference indicated that participants preferred a prospect more often when it was the competitor than when it was the target. No difference indicated that participants were consistent, and chose a prospect the same number of times regardless of whether it was the target or competitor.

Since in some trials participants chose the decoy option, and because these selections were positively correlated with target selections, we counted decoy selections as target selections in proportion to the degree to which they were correlated. This resulted in each decoy choice being coded as .67 in favour of the target and .33 in favour of the competitor. This method provides a stricter test of our hypothesis than apportioning decoy selections equally between competitor and target as was the case in the Wedell (1991) design. All
the following results were analysed using both decoy coding methods and were found to hold under either.

### 3.3.3 Results

Two participants were excluded from all analyses presented here due to non-compliance with the instructions. To check that participants perceived the decoy option as inferior to the target option we examined decoy selections for each of the 16 choice sets in both contexts. Participants sometimes chose the decoy option instead of the target or competitor. The overall decoy selection rate was 5.37%.

**Effect of Expected Value Difference**

Figure 3.6 shows that the preference reversal rate decreased as expected value difference increased. An expected value difference of zero yielded the highest preference reversal rate, while expected value differences of six and nine yielded preference reversal rates close to zero. A Friedman’s analysis by ranks gave a significant effect of expected value difference on preference reversal rate ($\chi^2(3) = 13.21, p = .004$). Post hoc analysis with Wilcoxon Signed Ranks Test was conducted with a Bonferroni correction resulting in a significance level $p < .008$. There was a significant difference between the zero and six expected value difference conditions ($p = .007$).

Figure 3.6: Experiment 1 results. Preference reversal rate for each level of absolute difference in expected value difference. Error bars are 95% confidence intervals.
3.3.4 Discussion

The results from Experiment 1 show that, in line with the predictions of the model, increasing the difference in expected value between target and competitor prospects reduces the effect of having an asymmetrically dominated decoy in the choice set. This suggests that as one option becomes clearly preferable to the other, people are less likely to be swayed by the presence of a decoy option, instead consistently choosing the option that they prefer. These data provide evidence that the attraction effect occurs in proportion to the uncertainty in determining which alternative is best.

In Experiment 2 we replicate the effect in two additional paradigms including the judgement of rectangle area, similar to Trueblood et al. (2013). The judgement of rectangle area is hypothesised to have a smaller variance than the calculation of expected value in prospects. Thus we predict that the attraction effect should subside more quickly in the area judgement tasks compared to the choices between prospects.

3.4 Experiment 2

3.4.1 Methods

We designed Experiment 2 to be a replication of Experiment 1 using a relative, rather than absolute measure of expected value difference. We also extended the design to include two further stimuli presentation types. This allowed us to replicate our findings and show that they extend to other paradigms that have been used to elicit the attraction effect. See stimuli section for more details.

Participants

One hundred and forty-four undergraduate Psychology students from the University of Manchester volunteered to take part, fifty in Experiment 2a, 52 in Experiment 2b and 42 in Experiment 2c. Participants received course credits for participating.

Design

In each trial participants chose between a target, competitor and decoy prospect. Our independent variable was the difference in expected value between the target and competitor prospects. Expected value difference had four levels 0, 20, 100 and 300%, reflecting the percentage increase in expected value from the smaller to the larger expected value prospect. Sixteen prospect pairs (See table 3.2) were created spanning probability and value space, four for each of the IV levels.

Each prospect pair was presented to the participants eight times with one prospect as target, and eight times with the other prospect as target. The decoys had either 20% fewer
Table 3.2: Stimuli values used in Experiment 2

<table>
<thead>
<tr>
<th>Reference prospect</th>
<th>Alternative prospect</th>
<th>Δ0%</th>
<th>Δ20%</th>
<th>Δ100%</th>
<th>Δ300%</th>
</tr>
</thead>
<tbody>
<tr>
<td>.12(83)</td>
<td>.24(42)</td>
<td>.29(42)</td>
<td>.48(42)</td>
<td>.96(42)</td>
<td></td>
</tr>
<tr>
<td>.17(59)</td>
<td>.24(42)</td>
<td>.29(42)</td>
<td>.48(42)</td>
<td>.96(42)</td>
<td></td>
</tr>
<tr>
<td>.59(17)</td>
<td>.42(24)</td>
<td>.42(29)</td>
<td>.42(48)</td>
<td>.42(96)</td>
<td></td>
</tr>
<tr>
<td>.83(12)</td>
<td>.42(24)</td>
<td>.42(29)</td>
<td>.42(48)</td>
<td>.42(96)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.7: Stimuli used in Experiment 2. The bottom panel shows the descriptive stimuli from Experiment 2b, while the top panel shows the rectangle stimuli in Experiment 2c.

value points, or 20% fewer probability points than the target. The experiment consisted of 256 trials (4 IV levels x 4 prospect pairs x 2 decoy positions x 8 repetitions) which were presented in random order. The dependent variable of preference reversal rate was calculated in the same way as in Experiment 1. See the Experiment 1 analysis section for details.

Stimuli

Experiment 2a, was very similar to Experiment 1, conveying the probabilities and values via the same grid display. The stimuli differed in that the probability grid was not randomised, there was no time limit and participants could view all three prospects simultaneously.

Experiment 2b presented the stimuli in a manner more similar to the original Wedell (1991) design, sentences were displayed on screen to convey the objective probabilities and values of each prospect (see the left panel in Figure 3.7). Prospects were presented simultaneously in sentence form ‘p probability of v points’.

In Experiment 2c we modelled our stimuli on those used by (Trueblood et al., 2013)
using rectangle area instead of expected value (see the right panel in Figure 3.7). Participants were presented with three rectangles and asked to choose the rectangle with the largest area. The height and width of the rectangles in pixels were the same as probabilities and values used for the prospects in 2a and 2b. Since both the area of a rectangle and expected value of our prospects are given by the product of their attributes, it was simple to substitute expected value difference for area difference.

**Procedure**

For all three experiments, participants were presented with three prospects or rectangles simultaneously on a computer monitor. Below each option a button marked ‘select’ allowed participants to indicate which prospect they preferred or rectangle they perceived to have the largest area. The on-screen order of the three stimuli in each trial was randomised. 256 trials were presented in four blocks of 64 trials with an enforced one-minute break between blocks.

**3.4.2 Results**

**2a**

The preference reversal rate was similar in the 0% EVD and 20% EVD conditions, but fell in the 100% and 300% EVD conditions. See figure 3.8a. The effect of expected value difference on preference reversal rate was assessed using Friedman’s analysis by ranks, this approached significance ($\chi^2(3) = 7.415, p = .06$).
The preference reversal rate fell steadily from the 0% EVD condition to close to zero in the 300% EVD condition. See Figure 3.8b. A Friedman’s analysis by ranks gave a significant effect of expected value difference on preference reversal rate ($\chi^2(3) = 20.57, p = .0001$). Post hoc analysis with Wilcoxon Signed Ranks Test was conducted with a Bonferroni correction resulting in a significance level $p < .008$. There was a significant difference between the 0% and 100% conditions, the 0% and 300% conditions and the 20% and 300% conditions (all $p =< .008$).

The preference reversal rate was low, around 5% in the 0 and 20% EVD conditions, falling to around zero in the 100 and 300% EVD conditions. See Figure 3.8c. A Friedman’s analysis by ranks gave a significant effect of expected value difference on preference reversal rate ($\chi^2(3) = 15.58, p = .001$). Post hoc analysis with Wilcoxon Signed Ranks Test was conducted with a Bonferroni correction resulting in a significance level $p < .008$. There was a significant difference between the 20% and 100% conditions and the 20% and 300% conditions (both $p =< .008$).

**Overall Results**

In agreement with Experiment 1, the results show that across different modalities, the attraction effect was sensitive to the expected value manipulation. As the difference in expected value between the target and competitor increased so the attraction effect declined. In all cases preference reversals were lower in the 100 and 300% expected value difference conditions than in the 0 and 20% conditions. Although there were differences between the preference reversal rates when different stimuli types were used, the general result was preserved across experiments, suggesting that as expected value difference increases the attraction effect is significantly reduced.

### 3.5 Discussion

We tested whether the attraction effect is moderated by the difference in expected value between target and competitor prospects, and for consistency with our model. We found that across several classes of experimental stimuli, including lower level perceptual judgements and higher level decisions from description, the preference reversal effect was greatly reduced as the difference in expected value increased. In short, we have provided an empirical demonstration that the attraction effect is moderated by whether or not participants can tell the difference between the expected value of alternative prospects. If people perceive a sufficient difference then they do not exhibit the attraction effect.
Our data provide evidence for an alternative interpretation of the attraction effect. Firstly, the observation that the effect violates axioms necessary to maximise value should carry the caveat that this only occurs when it is difficult to tell the utilities apart, peaking precisely at the point where choosing either alternative would maximise value anyway. Secondly we argue that to exhibit the attraction effect is actually an adaptive response to noise in determining value of each alternative. We expand on this argument below.

3.5.1 The Attraction Effect as an Adaptive Behaviour

The data obtained are well explained by our proposed model. Recall that an increase in expected value difference between target and competitor prospects was predicted to result in a correspondingly smaller attraction effect. This occurs in the model because the increase in expected value resulting from being a target option can only impact on choice share if it alters the extent to which target and competitor expected values overlap. The larger the difference in expected value, the smaller the likelihood that a shift in the target distribution will cause it to overlap with the competitor distribution. Our results across all paradigms were consistent with this hypothesis.

In Experiment 2c (rectangle area) the attraction effect was already absent in the 100% expected value difference condition. We suggest that this reflects the fact that performing the expected value (area) calculation is much easier for participants in this task. They need only observe each rectangle as opposed to combining two separate attributes, which was necessary in the other paradigms.

In the model, an increased certainty in expected value calculation means that the attraction effect should be more sensitive to changes in expected value difference. The increased sensitivity occurs because the relatively higher kurtosis of the perceived expected value distributions means that a smaller difference in expected value is necessary to eliminate the overlap between target and competitor. The data from Experiment 2c are consistent with this prediction assuming that judging the difference in rectangle area is subject to less variance than determining the expected value of a gamble.

One aspect of our data is not well captured by our model. This is the size of the attraction effect in the rectangle stimuli when there is no expected value difference. Although our model predicts a quicker reduction in the attraction effect as a function of expected value difference it predicts the same size attraction effect when there is precisely zero difference in expected value. Our data show that at zero expected value difference the attraction effect was smaller in the rectangle stimuli. One possible explanation is that participants did not perceive the rectangles to have the same area in the zero expected value difference condition because of the vertical horizontal illusion (Finger & Spelt, 1947) in which lines appear longer when displayed vertically than horizontally. A more intuitive explanation is that the attraction effect is simply smaller (at zero difference) when people
are more certain in their expected value estimates, and that this is a subtlety not captured in our high-level model. The reason the model does not make this prediction is because an effect size calculation is not sensitive to variance when there is no difference between the means.

One further caveat is that our findings and model relate to prospects and rectangles. These stimuli are something of a special case in that the attributes on which they are described are commensurate, allowing us to analyse performance in terms of expected value. However, the attraction effect has been shown in choices that are described by attributes that are less easy to integrate such as food quality and driving time for restaurants (Huber et al., 1982). It is an open question whether these results would be reproduced in choices described on non-commensurate attributes, although it seems likely that it would be hard to bias someone toward a one-star restaurant that is a 30 minute drive away, from a five-star restaurant that is a 35 minute drive away.

Demonstrating the adaptive nature of our model would perhaps be more difficult given non-commensurate attributes. It is worth noting however, that our analysis is a consequence of the ordinal relations within each attribute predicting that the target will be higher than average on one attribute while being average on the other. The competitor will also be higher than average on one attribute, but lower than average on the other. For most plausibly shaped indifference curves the target should be expected to be preferred.

3.5.2 Theoretical Implications

An unambiguous finding from our experiments is that the attraction effect will not be exhibited, given a sufficient difference in expected value between the alternative prospects. This suggests an important caveat for any claims about the attraction effect and discussion of value maximising models. The violations of these models are confined to scenarios where the alternatives are difficult to tell apart, implicit in this is that there may be little consequence to the decision since either alternative is likely to achieve near maximum value. This finding addresses important questions raised regarding preference reversals and violations of axioms in value-maximising models. In economics Roth (1996) has previously argued that the importance of preference reversals should be tested with expected value manipulations. In the cognitive sciences, Anderson (1990) has argued that violations of axioms do not violate the adaptive principle of rationality unless there is some difference in adaptive value between the alternatives.

We can go a step further than the claim that the attraction effect is not irrational and argue instead that exhibiting the attraction effect is actually adaptive. When expected value maximisation leaves some ambiguity as to which alternative is best, the additional information supplied by the ordinal relations results in some subtle deviations from the value-maximising axioms, but not from value-maximisation itself. According to this model we
might predict that human decision-making is well described by value-maximising models except in the scenarios where there is still substantial uncertainty as to the best alternative. It is precisely in these situations that contextual information such as that provided by the ordinal relations comes in to play, demonstrating not an inability to apply the normative rules, but instead going above and beyond, attempting to eke out as much value as possible from an uncertain environment. It is interesting to note that many of the violations of normative decision making rules are elicited using alternatives that have very similar expected values. For instance, violations of the independence axiom in expected utility theory can be elicited, but the difference in expected value between the alternatives is typically barely perceptible (Jarvstad et al., 2014).

An interesting aspect of our model is that we show that an IIA and regularity violating agent can achieve a higher expected value than one that does not violate these axioms. For a noisy expected value maximiser then, the normative solution in Luce’s choice axiom must be incorrect. (Luce, 1959) in fact considered a scenario in which the decoy provides information:

In such a situation the third alternative is by no means irrelevant to the choice between x and y, since it indicates to some degree what the outcome will be, and so axiom 1 would not be expected to hold. Put another way, if the third alternative is also a discriminative stimulus for the state of the universe, then, by definition, it will not be irrelevant to the choice between the first two alternatives whose outcomes depend upon which state obtains. (Luce, 1959, p. 132)

Rather than classify human decision making as violating rational models, it can be instructive to consider whether the normative solution is well-matched to the problem faced by the decision maker. With only the assumption that people have noise in calculating expected value it can be shown that the normative solution is different and that behaviour can still be understood as adaptive and value-maximising.

Perceiving the attraction effect as adaptive is a different type of explanation to most existing models. The dominant view is that the effect is a consequence of limiting neural mechanisms (Roe et al., 2001; Soltani et al., 2012; Usher & McClelland, 2004) or can only be explained with descriptive models (e.g, Tversky & Simonson, 1993). In these explanations the change in choice proportions is seen as a side-effect of more general decision making system or one that is incapable of achieving the normative standard. In contrast we argue that the effect is the result of an effort to maximise value. Interpreting the attraction effect as adaptive reveals perhaps why the behaviour occurs, as opposed to describing how the behaviour occurs. Whilst the mechanisms behind the attraction effect are undoubtedly interesting, an explanation of why the effect occurs might lead to a broader, more parsimonious model of the effects of context on decision-making.
3.5.3 Conclusion

In this paper we have presented an empirical investigation of the attraction effect and expected value difference together with a high-level model of the attraction effect as adaptive. Further theoretical work is ongoing, extending the insight of this model to other contextual effects and risk aversion (Howes, Warren, Farmer, El-Deredy, & Lewis, 2014).

The existing literature interprets the attraction effect primarily as a violation of a normative axiom. We argue that focusing instead on why the effect might be advantageous, allows us to pursue novel and parsimonious models of decision making such as understanding the impact of context as an adaptive response to noise inherent in decision making.

In situations where people can perceive the difference in expected value between prospects they do not exhibit the attraction effect and do not violate value-maximising models. When people choose between alternatives that are similar in expected value they do not fail to maximise value, they go above and beyond what an IIA consistent decision maker would achieve with the same noisy decision problem. We suggest that the attraction effect should be interpreted as an adaptive use of context to maximise value.
Chapter 4

Why Expected Value Maximising Organisms Exhibit Preference Reversals and Risk Preferences

Abstract

Preference reversals occur when a preference for one choice over another is reversed by the addition of further choices. It has been argued that the occurrence of preference reversals in humans suggests that they violate the axioms of rationality and cannot be explained with utility maximization theories. In this article we use numerical simulation to demonstrate that for a range of types of contextual preference reversal, including the attraction, compromise and similarity effects, the rational, expected value maximizing, choice between existing options is one that is influenced by the addition of new choices. The analysis assumes that people rationally integrate two sources of information, one based on an estimate of expected value and one based on observation of the order of feature values. We also show that the same assumptions explain some observed risk preference effects. We conclude that experiments showing that people exhibit contextual preference reversals and risk preference are not evidence that they are irrational; they are however evidence that they make expected value maximizing use of the available resources.

4.1 Introduction

One of the successes of the rational analysis of human decision making has been that a number of apparent irrationalities of choice have been shown to be rational given different assumptions about what it is that shapes adaptation (Oaksford & Chater, 1994; Hahn & Warren, 2009; Le Mens & Denrell, 2011). For example, Hahn and Warren (2009) have shown that a consideration of an individual’s experience of a fair coin toss, and of the bounds imposed by working memory, can explain seemingly biased perceptions of
randomness. Similarly, Oaksford and Chater (2007) proposed that, instead of questioning whether people are rational when compared with purely deductive logic, human behaviour might be understood as rational relative to the ecology of an uncertain world. Despite these successes one apparent irrationality that has so far resisted rational analysis is the contextual preference reversal (Huber et al., 1982; Huber & Puto, 1983; Wedell, 1991). In the current article, we report an analysis of contextual preference reversals that shows that human behaviour can be considered computationally rational (Lewis et al., 2014) given the uncertainty introduced by perceptual and cognitive capacities. We also show that risk preferences are an emergent property of this model. 1

Contextual preference reversals occur when a preference for one choice over another is reversed by the availability of further choices. Evidence that people make these reversals has been taken as evidence against a number of normative theories of human choice. In particular, preference reversal phenomena have been taken to suggest that human decision making cannot be explained by value maximizing theories of rational choice (Tversky & Simonson, 1993; Huber et al., 1982; Simonson, 1989; Ariely & Wallsten, 1995; Simonson & Tversky, 1992; Summerfield & Tsetsos, 2012; Louie & Glimcher, 2012; Tsetsos et al., 2010). The basis for this argument has been that preference reversals, in the presence of a new alternative, suggest that the values of each option are influenced by additions to the set of available choices, which are thought to be a violation of Luce’s Independence from Irrelevant Alternatives (IIA) axiom (Luce, 1977). As its name suggests the axiom demands that a preference between two options is not changed by the addition of a choice to the set of what is available. For some authors, preference reversals even suggest that organisms violate the axioms of rationality (Tversky & Simonson, 1993; Usher et al., 2008).

Consider a simple example of a preference reversal. When asked whether you would like a healthy apple A or a cake B that has 30g of sugar, you might choose the apple, but when subsequently told that the cake has less sugar than a second cake D, that has 40g of sugar, switch preference to B. A person who makes such a switch makes a contextual preference reversal. Typically, in preference reversal experiments choices are described in terms of two attribute dimensions and not just one (sugar). For example in the experiments, reported by Huber et al. (1982), the dimensions were, for example, financial cost and quality of a beer. In the Wedell (1991) experiments participants were provided with numeric probability and value attributes. In both of these circumstances, a choice between, say, A and B is difficult because while A might be higher than B on one dimension it will be lower on the other. The decoy D must be dominated by one choice, say,
B, but not by the other, on both dimensions; that is, D must have a worse value on both dimensions.

A gamble version of a preference reversal task is illustrated in Figure 4.1. Choice B has a higher value V than choice A but a lower probability P. Choice B dominates D but not A. Choices A and B have approximately equal expected value. In the Figure, given the position of B, A must appear in the shaded arc to the top left, and D in the shaded rectangle. In addition to multi-attribute decision problems (Huber et al., 1982; Huber & Puto, 1983; Simonson, 1989; Simonson & Tversky, 1992) and gambles (Wedell, 1991; Herne, 1999) other paradigms have been used to study contextual reversals. One study used political candidates as stimuli (Pan et al., 1995). More generally, reversals have been observed and are influential in neuroscience (Louie & Glimcher, 2012; Soltani et al., 2012), psychology (Trueblood et al., 2013), and zoology (Schuck-Paim, Pompilio, & Kacelnik, 2004; Latty & Beekman, 2011) and economics (Loomes, 2005). The effect has also been shown in birds (Shafir et al., 2002; Bateson, Healy, & Hurly, 2003), bees (Shafir et al., 2002) and ameboid organisms (Latty & Beekman, 2011). In humans it can be attenuated by increasing blood glucose levels (Masicampo & Baumeister, 2008), elaborating the descriptions of attributes (Ratneshwar et al., 1987) or simply by increasing the difference in expected value between the two options (Farmer, unpublished manuscript).

While models of decision making based on axioms such as Independence of Irrelevant Alternatives appear to fail to explain decision making in humans, a number of theories of the underlying processes and mechanisms that lead to these phenomena have been more successful. According to one of these mechanistic theories, Range-Frequency theory, choice D in Figure 4.1 extends the range of values on the y-axis thereby making B’s loss to A appear smaller than if choice D were not present (Huber et al., 1982). Alternatively, in motivational accounts of the preference reversal effects it was proposed that participant’s desire to be able to justify their choice to the experimenter leads them to prefer a dominating choice (Simonson, 1989). The argument is that it is easier to justify choosing B because it dominates one of the other choices and A does not. A connectionist model of choice, Decision Field Theory (DFT), based on relative values, explained preference reversals as a consequence of computing values from differences. Choice B wins over A because it has a bigger relative value to the decoy D than does A (Busemeyer & Townsend, 1993; Roe et al., 2001). DFT explains preference reversals as a side effect of neural processes that accumulate information and inhibit alternatives as attention switches between the available choices. The Leaky Competing Accumulator model offers another account of the mechanisms that might produce reversals but, unlike DFT, incorporates some non-linear assumptions and assumptions from Prospect Theory (Usher et al., 2008; Usher & McClelland, 2001). Lastly, Vlaev, Chater, Stewart, and Brown (2011) pursue an account in which people have no stable internal scales of value and there is no reason, therefore, to expect them to exhibit consistent preferences.
Figure 4.1: In a typical experiment two choices A and B are each described in terms of two attributes, here probability $p$ and value $v$. Each of A and B has a higher value on one of the attributes and a lower value on the other. A has a similar expected value to B and might therefore be placed anywhere in the shaded arc to the top left. A decoy D is placed in the shaded rectangle to the left and below B, so that it is dominated by B but not by A. The introduction of D is used to test whether there is an effect of this new alternative on the preference for A or B.

Despite the number and diversity of these process theories, none have attempted to explain preference reversals as a rational adaptation to the combination of bounds imposed by the task environment and the organism’s information processing capacity. In fact, many researchers have assumed that contextual preference reversals are evidence that people depart from the principles of rational choice. This belief has survived for at least 20 years. According to Tversky and Simonson (1993, p. 1179) “The standard theory of choice—based on value maximization—associates with each option a real value such that, given an offered set, the decision maker chooses the option with the highest value. Despite its simplicity and intuitive appeal, there is a growing body of data that is inconsistent with this theory.” In the same article, Tversky and Simonson (1993, p. 1188) state, “The analysis of context effects, in perception as well as in choice, provides numerous examples in which people err by complicating rather than by simplifying the task; they often perform unnecessary computations and attend to irrelevant aspects of the situation under study.” More recently, Usher et al. (2008, p. 297) state, “… contextual reversal effects … demonstrate a limitation of rationality in choice preference.” Soltani et al. (2012, p. 1) write “Decoy effects can be considered an error in logical reasoning.”

In the current article, we argue that preference reversals are rational given the assump-
tion that expected value estimates are subject to calculation noise. In these circumstances
the additions to choice sets made in Huber et al. (1982); Huber and Puto (1983); Wedell
(1991) are relevant to choice. Our article is a challenge to three aspects of the existing
literature: (1) models that fail to make use of information about the order of feature values
do not explain changes in preference between choices and, therefore, do not explain the
data, (2) mechanistic models fit the data but focus on explaining how contextual prefer-
ences reversals might occur as a consequence of connectionist architecture (Usher et al.,
2008; Roe et al., 2001) or as a consequence of a rank-dependent choice comparison pro-
cess (Tsetsos, Chater, & Usher, 2012; Vlaev et al., 2011; Tversky & Simonson, 1993),
(3) some researchers claim that preference reversals indicate that people are irrational
(Tversky & Simonson, 1993; Usher et al., 2008).

Our argument is based on an analysis of the computational rationality (Lewis et al.,
2014) of choice. The key idea is that the addition of alternatives to a choice set, even
dominated alternatives, adds information that must inform the posterior estimate of ex-
pected value. It does so by supporting the retrieval of prior expectations of choice utilities
(Bordley, 1992). We also show that risk preferences (risk seeking and risk aversion) can
be rational given the distribution of probability and value in the environment.

The plan for the paper is as follows.

1. In the Background section we review the expected value, preference reversal, and
risk preference literature. In so doing we attempt to elaborate the details of the exist-
ing assumptions and provide a basis for the assumptions that motivate our analysis.

2. We present a model of choice between gambles that is based on an optimal inte-
gration of two estimates of each gamble’s expected value. One of these estimates
is based on an observation of the rank order of feature values. The model is of a
situation where a human is asked to choose between options represented with two
features, a probability and a value, as is required in the preference reversal paradigm
reported by Wedell (1991), and where the decoy’s features are constrained accord-
ing to the standard preference reversal paradigms (Section 4.3). The model com-
bines two uncertain estimates that a choice is expected value maximizing. The first
estimate is based on uncertain calculation of expected value. The second is based
on the likelihood of the observed feature ordering (the ordering of the magnitudes of
the presented probabilities and values) given each possible expected value ordering.
The model uses Bayesian inference to combine these two uncertain estimates and
thereby maximize the expected value of the decision given the available informa-
tion and the uncertainty introduced by the perceptual/cognitive mechanisms. The
assumption that there is uncertainty in perceptual and cognitive processes, and per-
haps uncertainty in preferences, is assumed to be uncontroversial (Loomes, 2005;
Faisal, Selen, & Wolpert, 2008; Seymour & McClure, 2008; Maloney & Mamas-

3. We demonstrate that the model does indeed maximize expected value given the limitations imposed by the perceptual/cognitive system.

4. We apply the computationally rational model to a previously reported empirical study of preference reversals (Wedell, 1991). In so doing we demonstrate that the model predicts a broad range of preference reversal phenomena (including attraction, compromise and similarity effects) (Section 4.4.2). Importantly, the model predicts preference reversal in the 4 conditions where it is expected and predicts its absence where it is not observed in humans.

5. We show that the same model also predicts a risk preference that is dependent on the mean of the value distribution. The model explains risk preference phenomena that have been used to justify the abandonment of expected value in favor of Subjective Expected Utility as a model of the human utility function (Savage, 1954; Johnson & Busemeyer, 2010). Our model demonstrates that risk preference can be a computationally rational response to the distribution of features in the environment and uncertainty in the organisms capacity to estimate expected value.

4.2 Background

4.2.1 Expected Value and Computational Rationality

In the decision sciences if a gamble is represented as a random variable $X$ with discrete outcome values $\{x_1, x_2, \ldots, x_n\}$ that occur with probabilities $\{p_1, p_2, \ldots, p_n\}$ then the expected value (EV) of $X$ is,

$$EV(X) = \sum_{i=1}^{n} p_i x_i$$

The optimal choice given more than one gamble is to select the gamble with highest EV according to this definition. However, in models of human decision making, EV is replaced by one or other equation that transforms the objective value represented by EV with one or other subjective value utility. For example, in Subjective Expected Utility (SEU) theory (Savage, 1954) objective values can be weighted with a power function $U(x) = x^\alpha$. Probabilities can also be weighted to represent subjective probabilities. The weights were introduced so that the utility function could describe human behaviours such as the following: (A) a certain outcome valued at $1$ million, and (B) an uncertain outcome with an 89% chance of $1$ million, a 10% chance of $5$ million, and a 1% chance of nothing. Expected value suggests choosing gamble (B) but most people choose gamble (A). A fact
that can be described by adjusting the weights in SEU so as to capture the intuition that there is diminishing marginal utility to higher values. In other theories of subject utility other adjustments are made. In Prospect Theory (Kahneman & Tversky, 1979) losses and gains are given different utility functions. These theories, along with others that include Regret Theory and Cumulative Prospect Theory, are reviewed by Johnson and Busemeyer (2010).

Despite the success, and intuitive appeal, of the various theories of human subjective utility at describing human decision making, in Section 4.3 our starting point for explaining human behaviour will be expected value. We show that, given uncontroversial assumptions about perceptual/cognitive information processing limits, computationally rational decisions can lead to some of the phenomena that are usually used to motivate the subject weighting of probability and value.

### 4.2.2 Preference Reversals

The first empirical demonstrations of attraction effect preference reversals were for human choice between consumer products (Huber et al., 1982; Huber & Puto, 1983; Simonson & Tversky, 1992; Simonson, 1989). Other studies have used different stimuli types: political candidates (Pan et al., 1995), gambles (Wedell, 1991), and gambles with payment (Herne, 1999). We focus in what follows on the human studies of preference reversals in consumer product choice and gamble experiments and subsequently on the explanations of these phenomena.

The attraction effect preference reversals, as illustrated in Figure 4.1, were first demonstrated experimentally when participants were asked to choose between gambles from various product categories including cars, restaurants, and beers (Huber et al., 1982). Each pair of gambles presented to participants shared two attributes such as price and quality. As described in the Introduction, two gambles, \( A \) and \( B \), were approximately equivalent such that each dominated the other on one of the attributes. Two weeks after choosing between \( A \) and \( B \) in each of the categories, 93 participants returned to answer the same choice problems but with the decoy \( D \) added. The result, aggregated across product categories, was a 9 point increase in percentage share for the gamble that dominated the new decoy. A follow up study (Huber & Puto, 1983) showed that the effect can also be achieved when the decoy is inferior to but not strictly dominated by one of the choices.

The attraction effect has also been shown in choices between gambles (Wedell, 1991). The stimuli used all had an expected value of approximately $10, but with different probabilities and values. For example, one problem was to choose between \((.83, \$12)\), \((.67, \$15)\) and \((.78, \$10)\). The last gamble in this set, the decoy, can be ruled out because both its \( P \) and its \( V \) are less than those in the first gamble. Choosing between the first two gambles is difficult because while one gamble dominates on \( P \), the other dominates on \( V \).
A paired task was to choose between (.83, $12), (.67, $15) and (.62, $13). The first two choices are the same as before but the new decoy is dominated by the second choice rather than the first. In this paradigm, preference reversal can be measured as the proportion of pairs on which the dominating choice is chosen in both variants. This study resulted in about 20% of choices exhibiting the attraction effect preference reversal.

There have been many attempts to explain the preference reversal in terms of the underlying psychological mechanisms. Tversky and Simonson (1993) proposed that preference reversals can be explained in terms of two psychological processes: a process that weights the effect of the background to the decision, and a comparison process that describes the effect of the local context. The local context might consist of a choice between three beers, and the background, experienced before the local context, might have included five or six other beers. The background process increases the value of choices in the local context if, for example, they have a price that is lower than prices of beers in the background. The local context process increases the value of choices that are better than proximal choices that are also in the local context, and it does so by summing the relative advantage that each choice has over other choices in the set (Tversky & Simonson, 1993, p. 1186). The model calculates the relative advantage that each choice has over other choices on each attribute. In the attraction effect, the dominating choice’s relative advantage over the decoy exceeds the relative advantage of the other choice over the decoy. Therefore when the non-decoy choices are compared only with each other, they will have equal choice probability, but when the decoy is included in close proximity to one of the choices, this choice will have a higher choice probability.

Another process model, Decision Field Theory (Busemeyer & Townsend, 1993), was developed to examine the deliberation process in preferential decision making. This neurocomputational model was then extended (Roe et al., 2001) to explain contextual preference reversals. It is a connectionist network that accumulates preferences for each choice over time as the decision maker’s attention switches, stochastically, between the different gambles and their attributes. It is argued that the details of this accumulation process reveal how the attraction effect comes about.

In Multi-alternative Decision Field Theory (MDFT), each choice inhibits other gambles (Roe et al., 2001). The strength of this inhibition is inversely proportional to the distance between the gambles in the attribute space. During the deliberation, the decoy gamble comes to have a negative valence. Since the decoy dominating gamble is nearer than the other choice, it receives a bigger boost from comparison and is more likely to be chosen. Roe et al. (2001, p. 388) state that attraction effect reversals naturally ensue from the extended theory.

Another neurocomputational model, the Leaky Competing Accumulator (LCA) (Usher & McClelland, 2001) is a model of perceptual choice and, as with DFT, it has been applied to the problem of explaining contextual preference reversals (Usher et al., 2008). In
the model, a deliberation process involves comparing the attribute values of each gamble with each other gamble. The calculated differences are then transformed into a preference state via a loss averse value function. Because the non-dominating choice suffers two large disadvantages (one to the decoy and one to the other choice, on the y axis in Figure 4.1), and the dominating choice only suffers one large disadvantage (to the other non-decoy choice on the x axis) the non-dominating choice accumulates less preference and is chosen less often. This explanation is similar to that of Tversky and Simonson’s (1993) context dependent preference. In fact LCA can be seen as a neurally plausible implementation of that model (Usher et al., 2008). In LCA the value function is influenced by loss aversion. Usher et al. (2008, p. 297) argue that their account of preference reversals “demonstrates a limitation of rationality in choice preference.” In other words it is a side effect of a system adapted to other, or more general, purposes.

Other models also derive the implications of neural information processing constraints. It is possible that relative estimates of value may be a consequence of adaptation of neuronal firing to optimise sensitivity across large ranges of value Seymour and McClure (2008). In computational neuroscience, Soltani et al. (2012) have proposed a model in which stimuli are normalized so as to be distinguishable by neurons that have a firing rate of between 0 and a few hundred spikes per second. Without normalisation neurons would not be able to represent the range of experienced values. Soltani et al. (2012) show that this neural constraint can lead to preference reversals. An important contribution of this work is that it takes known facts about constraints on the operation of neurons and works through their implications for choice behaviour.

In contrast to the neurocomputational models a model proposed by Vlaev et al. (2011) takes as a point of departure the assumption that there are no stable internal utility scales. According to Vlaev et al. (2011) stable internal utility scales, which are needed in models that demand Regularity and IIA, are not present in humans and instead context is all important in decision making. This type of model is reviewed and contrasted to other types of choice model by Vlaev et al. (2011). They propose that the models can be categorized into three types. Type I, value-first models, compute the value of each gamble independently and then pick the highest. Type I models include expected value theory and expected utility theory. In Type II models, the value of each choice is computed and the highest picked but the value of each choice is influenced by the context. Type II models include DFT and LCA. In Type III models, decision are made purely by comparing choices at the time of decision and making relative judgments. Models that take this approach include Decision by Sampling (Stewart et al., 2006). Decision by sampling assumes that people only make comparisons among gambles rather than first computing an absolute value of utility. In decision by sampling, attribute values are compared in a sequence of ordinal comparisons. The number of favorable comparisons is recorded and the model chooses the first gamble with a count of favorable comparisons that exceeds a
theshold (Stewart, 2009, p 9).

**Summary**

Contextual preference reversals have been taken as evidence that the decision making processes of a number of organisms, including people, do not correspond to key axioms present in theories of how to make rational choice. This observation has lead to a prevalent view that the attraction effect suggests that value maximization approaches cannot be used to explain behavior (Tversky & Simonson, 1993). Consequently, the attraction effect is taken to be indicative of an overall tendency for decision making to be context dependent.

Further, influential computational explanations of contextual preference reversals suggest the effect is as an artefact of cognitive-neural information processing mechanisms (Usher et al., 2008; Roe et al., 2001). Preference reversals, according to these explanations, are an outcome of a bounded system failing to generate the normatively rational solution. In contrast, in what follows, we argue that the decoy provides information that implies that the decoy dominating choice is more likely than the other choice to be the best choice. We demonstrate that it can be rational to switch preferences.

### 4.2.3 Risk Preference

Risk preferences are evident when a decision maker appears, for example, to systematically prefer a gamble that is less risky in the face of gambles with higher expected value. While the primary purpose of the current article is to report an analysis of the computational rationality of preference reversals, the model that we develop in Section 4.3 also makes predictions concerning risk preference. In addition, key empirical studies that demonstrate preference reversals, also demonstrate risk preferences (Wedell, 1991).

While the optimal choice in any decision problem should be to pick the option with the highest EV and where EV is an average of the values weighted by their probabilities, risk preferences suggest that the probabilities and values are also weighted. In their motivation for cumulative prospect theory, Kahneman and Tversky (1979) and Tversky and Kahneman (1992), provide an extensive list of risk preference effects. These include: reference-dependence, loss aversion, diminishing sensitivity, and probability weighting. While conducted over 20 years ago this work has remained influential and is foundational to much thinking in behavioural decision sciences (Baucells & Villasís, 2010; Abdellaoui, Bleichrodt, & Kammoun, 2013; Glöckner & Pachur, 2012; Budescu & Por, 2013; Peel, 2013; Zeisberger, Vrecko, & Langer, 2012). In the current article we are interested in a subset of the known risk preference effects. We focus here on loss aversion, risk seeking, and risk aversion.

**Loss aversion** refers to the fact that people are more sensitive to losses than to gains of the same magnitude. Kahneman and Tversky (1979) cite evidence for loss aversion in
Figure 4.2: Given a choice task, the theory (to the left of the dotted line) is that people make two uncertain observations of expected value; one is based on a calculation of EV from feature values, the other is based on the feature rank orders. The analysis of computational rationality (to the right of the dotted line) derives the optimal decision given this information. It does so by updating the probability of each of 7 hypotheses given the observations. The posterior probability is used to predict the best decision given the uncertainty.

The fact that most people turn down the gamble (−$100, 0.5; $110, 0.5). People tend to be risk-seeking over losses: they prefer a 50 percent chance of losing $1,000 to losing $500 for sure. People tend to be risk averse over the domain of gains: they typically prefer a certain gain of $500 to a 50 percent chance of $1,000. Many empirical studies have observed these preferences. Tversky and Kahneman (1992), for example, observed this pattern in a study of 25 participants. They presented participants with a series of choices between prospects with monetary outcomes and numerical probabilities. There were 28 positive and 28 negative prospects. The data were used to calculate a certainty equivalent and the proportion of participants who exhibited a risk preference. The results showed that for all probabilities greater than or equal to 0.5, all participants were predominantly risk averse for gains and risk seeking for losses.

4.3 A Model of Computationally Rational Choice

We assume that the decision problem in a preference reversal experiment is to choose the option \( i \) that will maximize the expected value (the best option) where options are sampled from a distribution of possible objects. Each object is represented by a set of features each
of which has a probability and a value.

For a three choice problem there are 7 mutually exclusive hypotheses. These include $H_A$, the hypothesis that $A$ is the best option, $H_B$ the hypothesis that $B$ is the best, $H_D$ the hypothesis that $D$ is the best. The remaining hypotheses concern the possibility that two or more of the choices have the same expected value. These hypotheses are: $H_{AB}$, $H_{AD}$, $H_{BD}$, and $H_{ABD}$. In the absence of any evidence about a particular choice task, the prior probability that any one of these hypotheses is true is the prior probability that each is true given the distribution of choices in the experienced environment.

We assume that participants make two independent observations of the evidence concerning each of these hypotheses, where the evidence is the information presented in the experiment. The first observation involves reading the features provided, e.g., a probability and value in the case of Wedell (1991), and making a mental calculation of the expected value $\{U_A, U_B, U_D\}$ of each choice. The observation of each expected value is noisy, particularly because of the difficulty of calculating expected value given feature information. Given these constraints we assume that the computationally rational observer updates the prior estimates of the probability of each hypothesis given the evidence provided by the utility calculation. The outcome is a posterior probability $P(H_j|U_A, U_B, U_D)$ of each hypothesis $j$ given the evidence. This update is represented by the processes illustrated in the top row of Figure 4.2.

The second observation involves an encoding of the feature orders $R$ with independent observation noise. The likelihood of this observation is then used to update the probability of each hypothesis to give a new posterior $P(H_j|R)$. The model uses the hypothesis $H_j$ with the highest $P(H_j|R)$ to determine the choice. If the hypothesis suggests that two or more choices have equal value then a random choice is made between those options.

A formal description of the model is given below.

1. For three choices $\{A,B,D\}$ there are 7 mutually exclusive and exhaustive hypotheses $H_j$ where

   \[
   H_A \iff U_A > U_B \cap U_A > U_D \tag{4.1}
   \]

   \[
   H_B \iff U_B > U_A \cap U_B > U_D \tag{4.2}
   \]

   \[
   H_D \iff U_D > U_A \cap U_D > U_B \tag{4.3}
   \]

   \[
   H_{AB} \iff U_A = U_B > U_D \tag{4.4}
   \]

   \[
   H_{AD} \iff U_A = U_D > U_B \tag{4.5}
   \]

   \[
   H_{BD} \iff U_B = U_D > U_A \tag{4.6}
   \]

   \[
   H_{ABD} \iff U_A = U_B = U_D \tag{4.7}
   \]

2. Each choice $i \in \{A,B,D\}$ has a set of $1..m$ feature probabilities $p_{i,1} \ldots p_{i,m}$ and each
of these probabilities has a paired feature value \( v_{i,1} \ldots v_{i,m} \). The probabilities are sampled from a probability distribution with range \([0; 1]\) and the values \( v \) are sampled from a probability distribution with a defined central tendency and spread. These distributions allow the modeling of the real-world distributions of probability and value in the environment. Here we assume that the probabilities are Beta distributed (with shape parameters \( a \) and \( b \)) and the values are Gaussian distributed (with parameters \( \mu \) and \( \sigma \)) but the model is not committed to these particular distributional assumptions.

\[
p \sim \mathcal{B}(a, b) \quad (4.8)
\]

\[
v \sim \mathcal{N}(\mu, \sigma^2_{\text{val}}) \quad (4.9)
\]

3. The first observation involves reading the numbers provided, a process which is subject to unbiased additive noise (\( \sigma^2_{\text{obs,}p}, \sigma^2_{\text{obs,}v} \)) and calculating the expected value of each choice. The observed expected value of \( U_i \) is the sum of the expected value given noisy observations of \( p \) and \( v \) and further noise \( \sigma^2_{\text{calc}} \) due to the calculation process.

\[
U_i = \sum_{k=1}^{n} (p_{i,k}v_{i,k}) + \phi
\]

\[
\phi \sim \mathcal{N}(0, \sigma^2_{\text{calc}}) \quad (4.11)
\]

4. \( M \) is defined as the ranking of the set of utilities \( U_A, U_B \) and \( U_D \) defined by Equation 4.10. Given the likelihood \( P(M|H_j) \) of observing choice utility ranking \( M \) (the probability of the data given the hypothesis), the posterior probabilities \( P(H_j|M) \) can be calculated with Bayes’ rule,

\[
P(H_j|M) = \frac{P(M|H_j)P(H_j)}{\sum_{i=1}^{7} P(M|H_i)P(H_i)} \quad (4.12)
\]

5. The second observation involves an independent encoding of a set \( R \) of feature \textit{orders}. We define a function \( f \) which, subject to a small probability of random re-ordering \( P(\text{error}_f) \), maps pairs of real numbers to the set \{\text{lessthan, equal, greaterthan}\}:
6. The likelihood of the observed ordering $R$ is used to make a Bayesian update of the posterior obtained from the utility calculation (Equation 4.12). In other words, given the observed feature ranking $R$, the posterior probability of hypothesis $H_j$ is calculated with Bayes’ rule,

$$P(H_j|R) = \frac{P(R|M,H_j)P(H_j|M)}{\sum_{i=1}^{7} P(R|M,H_i)P(H_i|M)}$$

(4.14)

Where the prior $P(H_j|M)$ is the posterior probability defined in Equation 4.12 and the denominator is the sum of the likelihood times the prior for each of the 7 hypotheses.

7. The posterior estimates of all 7 hypotheses $P(H_A|R)$, $P(H_B|R)$, $P(H_D|R)$, ..., $P(H_{ABD}|R)$ are compared and the largest determines the choice $\pi^*$.

$$\pi^* = \arg\max_{\pi} \pi \in \{P(H_A|R), P(H_B|R), \ldots, P(H_{ABD}|R)\}$$

(4.15)

4.4 Results

4.4.1 Expected Value Maximization

To test the claim that it is rational to make use of ordinal feature information, we randomly generated three-choice problems (tasks) from an environment in which options had a probability feature $p$ and a value feature $v$. The probabilities were Beta distributed with parameters $(a = 2, b = 2)$. The values were normally distributed with parameters $(\mu = 100, \sigma_{val} = 5)$. We defined a correct choice as the choice that maximized the true expected value. We then used a model based on the analysis above to make a choice between the three options and we recorded the proportion of correct choices made given 1 million Monte Carlo trials. We compared the proportion of correct choices for this computationally rational model (the posterior model) to the proportion of correct choices made by a model that only made use of expected value calculation (the utility model), and the choices made by a model that only made use of ordinal feature observations (the feature model). These comparisons are illustrated in Figure 4.3.
4.4.2 Preference Reversals

In this section we test whether the computationally rational choice model presented above generates the preference reversals observed in one well known preference reversal study (Wedell, 1991). We described this study briefly in Section 4.2 and expand this description here.
Figure 4.4: Choices A, B and D in one of the set of positions (R, F, RF, and R-prime) used by Wedell (1991) and extended for Compromise and Similarity positions.

In each of the experimental conditions there were 3 choices \( \{A, B, D\} \). In the A-decoy condition the decoy was dominated by choice A and in the B-decoy condition the decoy was dominated by B. Each participant made two choices for each pair of gambles. Choice A always had a higher probability and choice B a higher value.

In the first two experiments, four positions for the decoy relative to the target were tested (Wedell, 1991). These are illustrated in Figure 4.4. The Range decoy (R), Frequency decoy (F), Range-Frequency decoy (RF) and R’ (R-prime decoy) were each set to test a different hypothesis. A preference reversal effect was, as expected, observed in the R, F, and RF conditions. Also as expected, the effect was not observed in R-prime, where the decoy is positioned centrally between with choices A and B. Figure 4.4 also shows decoy positions for a Compromise condition (C) and a Similarity condition (S) that were not studied by Wedell et al. but which have been studied by others (Simonson, 1989; Trueblood et al., 2013). Trueblood et al. (2013) have demonstrated the presence of preference reversal effect for Compromise conditions, and of inverse reversals for Similarity conditions.

The probabilities and values used to model the R, F, RF and R-prime tasks were those used by (Wedell, 1991) with the exception that some of the probabilities were adjusted a small amount to ensure that all A and B choices had exactly the same expected value (which was not true of the original materials). Probabilities and values for the Compromise and Similarity conditions were extrapolated from Wedell et al.’s materials as described above.

We wanted to know whether preferences reversals would be predicted as a conse-
Figure 4.5: Reversal effects for model predictions and data from Wedell (1991). The table at the top is for human participants (C.I.s unavailable; Compromise and Similarity effects unavailable) and the bottom table is the contingencies predicted by the model (95% multinomial C.I.s). The model was fitted to the inverse and decoy rates, but not to the reversal rate and reversal rates are therefore predictions.

sequence of uncertainty in expected value calculation. For this reason we made the following parameter assumptions:

- **The environment.** The mean and standard deviation of the environment distribution for feature values $v$ were set to the predictive mean and standard deviation of Wedell’s task distributions (a scaled, shifted $t$-distribution with $d.f. = 9$). The shape parameters $(a,b)$ of the Beta distributed probabilities $p$ were set to the maximum likelihood values given Wedell’s task distributions.

- **The observation noise.** The observation standard deviations, $\sigma_{obs,v}$ and $\sigma_{obs,p}$, were set to 0.

- **Probability of a feature order error** was set to 0.

- **Utility calculation noise** was fitted to the inverse-reversal rate in the four conditions reported by Wedell (1991). An inverse-reversal occurred when participants favored the choice that did not dominate the decoy. The fitted value of utility calculation noise was $\sigma_{calc} = 0.2$. The sum of squared errors was 0.0169.

We contrast the preference reversal predictions of the fitted model to the human data in Figure 4.5. The top table of the figure shows the human reversals and inverse reversals
for each of the Range (R), Frequency (F), Range-Frequency (RF), and R-prime (R’). It shows that people exhibited more reversals than inverse reversals in all conditions except R-prime. Wedell (1991) reports that these results were significant. The model effects are shown in the bottom table of Figure 4.5. As described above, the utility calculation error was adjusted to fit the model’s inverse reversal rates to the human data. The reversal rates were not fitted but are predictions given the fit. The model predicts preference reversals in R, F and RF conditions. The predicted preference reversal rate is consistently and substantially higher than the inverse reversal rate in these conditions. As expected, the model predicts no effect of R-prime (Wedell, 1991). In addition, the model predicts an effect of Compromise (C), which was not studied by Wedell, and predicts a small inverse effect of Similarity (S), which is consistent with the human performance in direction if not in magnitude (Simonson, 1989; Soltani et al., 2012; Trueblood et al., 2013).

The decoy was selected by Wedell’s participants on 2% of R, F, RF and R-prime trials. It was selected on 0% of the fitted model’s trials in these conditions and in the Compromise condition. Further inspection of the model’s prediction of the Similarity effect showed that the decoy was selected on 73% of trials.

Figure 4.6 shows how the model’s preference reversal predictions are moderated by changes to the parameter values. In all panels, the parameters, other than the one that was changed, were set to the following values: \( \{t_{val}(9), Beta(a = 1, b = 1), \sigma^2_{obs,p} = 0, \sigma^2_{obs,v} = 0, \sigma^2_{calc} = 0.2, P(error_f) = 0\} \).

- **top left panel.** Here R, F, RF, and Compromise show a preference reversal effect irrespective of variation in \( v \). There is no effect of R-prime and there is a small inverse effect of Similarity.

- **bottom left panel.** Here the preference reversal effects are somewhat attenuated by decreasing variation in the probability distribution representing this feature in the task environment. The Compromise effect disappears completely when the Beta distribution parameters \( a = b = 5 \).

- **top middle panel.** Here increased noise in the observation of values attenuates the preference reversal effect for R, F, RF and Compromise. In contrast, while there is no preference reversal effect in the Similarity condition at \( \sigma_{obs,v} = 0 \), the effect is present at higher values of this parameter. The negative effect of higher values of this parameter on preference reversals is explained by the fact that the parameter has consequences both for the acuity of the utility estimate based on calculation and the acuity of the estimate derived from feature ordering.

- **bottom middle panel.** Here increased noise in the observation of probabilities also attenuates the preference reversal effects and introduces a Similarity effect.
Figure 4.6: Reversals minus inverse reversals against levels of each of six model parameters. The arrangement of the panels correspond approximately to the arrangement of the model components specified in Figure 4.2. The two left most panels represent the effect of the environment distribution on preference reversals. The middle panels represent the effect of observation noise. The right panels represent the effect of utility calculation noise and the probability of an order error.

- **top right panel.** The preference reversal effect is present and stable as long as the utility calculation noise is non-zero. The abrupt change from no effect at $\sigma^2_{\text{calc}} = 0$ to a substantial effect at $\sigma^2_{\text{calc}} > 0$ is due to the fact that in Wedell (1991)’s materials the non-decoy choices had the same expected value.

- **bottom right panel.** Again, the preference reversal effect is attenuated by increase levels of noise, here due to an increased probability of a feature ordering error.

It is evident from the above exploration of the parameter space that positive, non-zero, utility calculation error and lower levels of observation noise and feature ordering error are required to observe preference reversals.

**Discussion**

The analysis above demonstrates that the computationally rational preference model presented in Section 4.3 can predict human performance on Wedell’s (1991) choice tasks.
Before considering the implications of the model for risk preference, we briefly generalise the model to a simple two choice situations. This generalisation provides a more compact description of some of the key assumptions. We assume that the decision maker has no a priori knowledge about the utility of presented choices. In other words the decision maker is maximally uncertain about the utility of the choices that are offered. We model this situation with two choices, $A$ and $B$, with utilities that are random variables, $U_A$ and $U_B$, sampled from the same distribution. Assuming $U_A \neq U_B$ and in the absence of any other information $p(U_A > U_B) = \frac{1}{2}$.

Given the addition of a third random variable $U_D$, also sampled from the same distribution, where $U_D \neq U_A$ and a dominance constraint $K = U_B > U_D$, then by a simple application of Bayes’ rule we see that $p(U_B > U_A|K) = p(U_B > U_A \text{ and } K) / p(K)$. The numerator and denominator can be calculated simply by listing all 6 possible and equally likely dominance relationships between the 3 values. In 3 of these scenarios, namely $(U_B > U_A > U_D), (U_B > U_D > U_A)$ and $(U_A > U_B > U_D)$ the constraint $K$ holds. Consequently $p(K)$ is $\frac{3}{6}$. However, in only 2 dominance relationships, namely $(U_B > U_A > U_D)$ and $(U_B > U_D > U_A)$, does both $K$ and $U_B > U_A$ hold. Consequently the numerator $p(U_B > U_A \text{ and } K) = \frac{2}{6}$ and it follows that $p(U_B > U_A|K) = \frac{2}{3}$.

Consequently, given a choice between random variables $U_A$, $U_B$ and $U_D$, it will be optimal to prefer $U_B$ over $U_A$ given only the information that $U_B > U_D$. In other words, if $U_B$ dominates $U_D$ then $U_B$ should be preferred over $U_A$. This analysis shows that it is rational for a preference ordering between two choices to be influenced by information about the relative value between one of these choices and a third choice. This analysis holds as long as there is some uncertainty about the value of the choices.

### 4.4.3 Risk Preference

We examined whether the model, with parameter values fitted to the preference reversal effect (the values reported above), also predicted the risk preference observed by Wedell (1991). Figure 4.7 contrasts the rational risk preference, predicted by the model (bottom panel), to the observed risk preference (top panel). While the model predicts a smaller risk preference than was exhibited by participants, the direction of the predicted effect is the same as for the participants in all four of the conditions studied by Wedell et al.

So as to further understand the consequences of the model assumptions for risk preference we explored the parameter space. Figure 4.8 consists of 6 panels each of which shows the effect of one of the 6 noise parameters. In the top left panel, as the variance of the distribution of $v$’s is manipulated, the model always predicts risk averse behaviour in all conditions except Similarity. In the bottom left panel, the spread of the environmental probability distribution moderates the risk preference in all conditions. More uniformly distributed probabilities (low values of Beta shape parameters) tend to result in more risk
Figure 4.7: The effect of condition on risk preference. Top panel: human data from Wedell (1991) (C.I.s unavailable; Compromise and Similarity effects unavailable). Bottom panel: The model risk preference. Values from a model using the fitted parameter values derived in the analysis of preference reversals (95% multinomial C.I.s).

aversion behaviour. In the top middle panel, increased value observation noise tends to increase risk aversion. In the bottom middle panel, increased probability observation noise tends to increase risk seeking. Conversely, the more accurately that probabilities are perceived then the more risk averse the choices. In the top right panel, utility calculation noise increases risk aversion in all conditions except Similarity. In the bottom right panel, higher rates of feature ordering error tend to lead to more risk neutral choice preferences.

While the above analysis suggests that the computationally rational model, based on Expected Value, can capture both risk averse and risk seeking behaviour, the results provide only a preliminary suggestion of the implications of the model. These are explored further in Figure 4.9. The figure presents five panels illustrating the effect of the ecological distribution of probability and value on the utility of different risk preferences for two option choice problems defined by the following sampling distributions: \( v \sim N(x, 30) \), \( p \sim B(1, 1) \). In the top left panel of the figure, \( EV_{max} \) is the optimal strategy; \( Risk\ averse \) is a strategy that always chooses the highest probability option; \( Risk\ seeking \) is a strategy that always selects the lowest probability option; \( Value\ seeking \) is a strategy that always selects the highest value option, regardless of its probability. Which of the strategies is best, if \( EV_{max} \) is unavailable, depends on the mean of the distribution of value \( v \). In the domain of losses, when \( mean(v) < -50 \) the risk seeking strategy approaches the optimal strategy. It approaches 100% efficiency (0 loss) as \( mean(v) \) de-
creases. In contrast, when mean(v) > 50 the risk averse strategy is closest to optimal, approaching EV as mean(v) increases. Lastly, when mean(v) \sim 0 and there are as many gains as losses then value seeking is closest to the optimum. Further, the value seeking strategy will exhibit loss aversion as it prefers options with smaller losses irrespective of probability.

The differences in the efficiency of risk seeking, risk aversion and value seeking, that are exhibited in the top left panel of Figure 4.9, are explained by the other panels of the figure. In the top right panel, density of EV is plotted for 5 means of the v distribution. While with mean(v) = 0 the distribution is symmetric, the distribution is increasingly skewed with increasing or decreasing mean; but more important than the skew is the fact that high probability and low probability options are not evenly distributed within these densities. This is shown by the bottom panels. In the left bottom panel, in the domain of losses, options with p < 0.5 have, on average, a higher EV than options with p > 0.5 which explains why it is better to be risk seeking in the domain of losses. In contrast, in the right bottom panel, in the domain of gains, options with p < 0.5 have, on average,
Figure 4.9: Five panels illustrating the effect of the ecological distribution of probability and value on the Expected Value (EV) of different risk preferences for two option choice problems defined by the following sampling distributions: \( v \sim N(x, 30), p \sim B(1, 1) \). **Top left panel:** A plot of the Loss in EV (measured as average points less than EV max) of 3 choice strategies. **Top right panel:** A plot of the EV density for 5 values of \( x \). **Bottom panels:** Plots of EV density for options with \( p < 0.5 \) and \( p > 0.5 \) (in the domain of losses, neutral, and gains respectively).

According to the above analysis, in the absence of reliable EV information, risk preferences are, on average, computationally rational. A particular preference will not always lead to the choice of the option with the highest EV but over the long-run, the best preference for the domain will more closely approximate the optimum than the other preferences.

### 4.5 General Discussion

We have presented an analysis of choice that demonstrates the rationality of preference reversals. The analysis optimally integrates two independent estimates of the probability.
that each choice has the highest expected value. Two sources of estimates were considered, the first was based on calculation and the second was based on the likelihoods of the observed feature rank ordering. Preference reversals were predicted when the EV calculation was uncertain. In these circumstances the perceived feature rank ordering influenced choice leading to more preference reversals. This analysis was shown to predict the various preference reversal effects, including Attraction and Compromise effects, given a model fitted only by fitting the model's EV uncertainty to a measure of the uncertainty in the participant's behaviour. In addition, despite being based on EV, the same model predicted the observed risk aversion, risk seeking and loss aversion without recourse to a parameter for weighting probability, or value, information.

The main implication of the analysis is that an organism that, subject to uncertainty, makes use of information about the rank order of feature values (dominance relations) will make better decisions (higher expected value) than one that does not. If an organism that preference reverses in the presence of a dominated choice does so in environments where the choice matters, then its fitness (propensity to utility maximize) will be greater than if it does not. Preference reversals are not irrational, they are not illogical, they are not a departure from the axioms of rationality as they should be applied to understanding the behaviour of computational organisms. On the contrary an organism that fails to preference reverse will fail to gain the expected value implied by the information in the task. Similarly, an organism that makes use of information about the distribution of probabilities in the environment may be an organism that will make better decisions (higher expected value) than one that does not. While we have provided only a preliminary analysis of risk preference, this analysis shows that an organism, should exhibit a risk preference, either risk aversion or risk seeking, depending on this distribution of probabilities.

Similarly an organism that is sensitive to the density of expected value in the environment, conditioned on the probability of the option, and subject to uncertainty imposed by the perceptual/cognitive systems, will make better decisions than one that does not. To be clear, in the absence of the ability to calculate EV, an EV maximizing organism is rational if it is risk seeking in the domain of gains, risk averse in the domain of losses and loss averse in mixed domains. As a consequence, there is no need to abandon EV as a theory of subjective utility in order to capture the basic risk preference effects.

One potential limitation of the analysis is that it says only a little about the underlying information processing mechanisms. The extent to which rational analysis can inform theories of mechanism has been the subject of a recent debate (Bowers & Davis, 2012; Griffiths, Chater, Norris, & Pouget, 2012; McClelland et al., 2010; Norris & Kinoshita, 2010). From one point of view our analysis says only that a rational information processing mechanism should make preference reversals when dominated choices are added to the context. This is a contribution at the level of computational theory (Marr, 1982; Oaksford & Chater, 2007). It is made by studying the environment of cognition, as rec-
ommended by Anderson (1990). However, the analyses say more. By virtue of the fact that decisions are bounded by limitations on an organism’s ability to calculate expected value, the analysis reveals the rational choice for a bounded organism. In this sense, the analysis shows that preference reversals are *computationally rational* (Lewis et al., 2014; Howes et al., 2009). Specifically, the analysis suggests that under uncertainty induced by the inevitable limitations of biological information processing systems, a mechanism can improve the utility of choice if it makes use of information about feature rank order.

Importantly, our analysis does not show that preference reversing organisms integrate multiple estimates of choice ordering optimally, i.e., it does not show that organisms use unbounded Bayesian algorithms to compute choice. Equally, the extent to which people are, or are not, Bayesians has no bearing on the implications of the analysis. Instead, Bayesian inference is used in our analysis merely as a tool for calculating the implications of the hypothesised information processing mechanisms and limitations. We hypothesised that preference reversals were a consequence of an information processing mechanism that combined uncertain expected value estimates with estimates of which choice was best based on feature rank ordering (Figure 4.2). The implications of these mechanisms for rational choice were then derived using Bayesian inference.

What our analysis does show is that making use of dominance information is rational and leads to preference reversals. There are many less rewarding algorithms for integrating information from multiple sources that would not be optimal but that would generate higher expected value than either expected value calculation or feature ordering alone. A bounded organism that could only compute approximations to a Bayesian integration should still exhibit computationally rational preference reversals so long as it had some capacity to combine rank dependent and expected value estimates. What our analysis does show is that, in these circumstances, the integration of calculative and feature rank ordering information leads to preference reversals and as a consequence to better decisions.

An implication of this observation is that a wide range of theories of the cognitive mechanisms that predict preference reversals may be rationally adapted to the choice task. DFT, LCA, and Range-Normalization, which offer process explanations of *how* contextual reversals might arise as a consequence of interactions between units in a parallel distributed network (Roe et al., 2001; Usher et al., 2008) or interactions between neurons (Soltani et al., 2012), are theories of what is rational given processing limitations. Further, it is conceivable that the rank dependent mechanisms proposed and reviewed in Tversky and Simonson (1993); Roe et al. (2001); Usher et al. (2008); Tsetsos et al. (2012), or the comparison only models (Vlaev et al., 2011; Stewart et al., 2006), are neurally plausible mechanisms by which contextual priors can be taken into account. In other words, making comparisons between choices, by whichever means, may just be an efficient way for a bounded organism to deal with uncertainty. What the analysis in the current article shows
is that, contrary to some views, these mechanisms and strategies may generate rational rather than distorted preferences under uncertainty.

The analysis of computational rationality that we have used adopted a methodological optimal approach, as advocated by Oaksford and Chater (1994). While the models make use of optimization to gain their explanatory force, they do not demand that people perform optimizations nor that they maximize utility to the extent shown to be possible by the particular Bayesian account above. In fact, it is useful to distinguish between a strong claim and a weak claim. The strong claim is that organisms optimally integrate feature rank ordering estimates of EV. The weak claim is that organisms make some use of feature rank ordering when expected value calculation is uncertain and that doing so is rational. By virtue of the fact that they demonstrate preference reversals, the data support the weak claim. The strong claim may certainly be interesting to pursue but to our knowledge the available contextual preference reversal data do not help ascertain its validity.

Further, our claim as to the rationality of the analysis is not based on its correspondence to a mathematical norm. The analysis is not rational because it uses Bayesian inference. Rather, the claim is made on the basis that the Value generated by the model’s behaviour is optimal given the available information (Russell & Subramanian, 1995; Howes et al., 2009; Lewis et al., 2014). This definition of rationality, we have argued, offers a way for rational analysis (computational rationality) to inform theories of biological mechanisms. In the proposed model the available information include uncertain expected value calculations and feature rank orders, but it might contain other sources of information about value. For example, in unpublished work we have shown that preference reversals are also a consequence of an agent that can estimate difference between feature values.

The prediction of risk preference effects from an analysis of expected value, in the absence of the probability and value weighting parameters dedicated to modeling risk preference that is found in other models of utility (Savage, 1954; Kahneman & Tversky, 1979; Tversky & Kahneman, 1992; Johnson & Busemeyer, 2010), provides further evidence in favour of the model. The model predicted the risk aversion effect observed by Wedell (1991) and offers an explanation of risk preference in terms of the variation in the environmental distribution of outcome probability. For half a century (Savage, 1954) risk preference has been modelled as a weighting of probability. Though much work is still required, the outcome of our analysis suggests that a deeper explanation of risk preference may be close; one that does not take the path of describing risk preferences with free parameters but rather derives what is expected value maximizing given plausible features of the adaptation environment and organism. If this is the case then it may be possible to explain variations in risk preference with task, as for example reported by Wu et al. (2009), in terms of variations in the structure of the environment or, at least, the perceived environment.
Our focus on what is rational given the constraints imposed by cognitive mechanisms is one difference with the approach of Shenoy and Yu (2013) who also describe a Bayesian model of contextual choice. Where we propose that preference reversals are rational given the limitations of the human cognitive architecture, Shenoy and Yu (2013) propose a normative account, following Marr (1982)’s framework, and informed by economic theory, in which the assumptions concern what people believe is fair in the marketplace. Further, where we focus on the integration of multiple sources of information given uncertainty in true EV and uncertainty in expected value estimation, Shenoy and Yu (2013) assume that uncertainty in posterior beliefs about market conditions contributes to randomness in choice on repeated presentations of the same options. The essence of Shenoy and Yu (2013) explanation is that the introduction of a decoy moves “the indifference line”, which is an estimate of fair value in a perceived market. The essence of our explanation is that introduction of decoy gives the decision maker the opportunity to make improved estimates of posterior expected values. Also, an important difference is our derivation that 

\[ p(U_B > U_A | K) = \frac{2}{3} \]  

given that \( K = U_B > U_D \) which links the contextual preference reversal phenomena to other decision making behaviours that can be explained with an analysis of the Monty Hall problem, e.g. see (Krauss & Wang, 2003). This analysis, Section 4.4.2 provides a general demonstration of the rationality of preference reversals in the absence of observation noise.

Before ending, what can we now say about Tversky and Simonson (1993)’s, and others, rejection of people as value maximizers? We have shown that the choices predicted by our analysis are value maximizing given uncertainty in true value and given the uncertainty introduced by the estimation and calculation processes. Our analysis makes the simple assumption, uncontroversial in statistical decision theory, that a value maximizing agent working under uncertainty should make use of the available information and Bayesian update to calculate the expected values of choice. Value maximization requires making best use of all of the information available according to its precision, and the results show that the more reliable the estimates from a particular source then the more positively they are weighted. Our analysis not only shows that contextual preference reversals should not be taken as evidence against value maximizing in people, rather in addition, it shows that preference reversals in human behaviour should, in fact, be read as positive evidence that people may be value maximizing given the limitations of their neural mechanisms, i.e., they may be computationally rational.

In conclusion, we claim that the analyses presented above are evidence that contextual preference reversals are not, as has been thought by many, evidence that people are “irrational” or that they “err,” or that they are not “logical.” Rather, it appears that these preference reversals are a consequence of the value maximizing use of contextual information by an organism operating under uncertainty.
Chapter 5

The Attraction Effect is Adaptive for Gains but not Losses

Abstract

The attraction effect occurs when preferences between alternatives are altered by the presence of an irrelevant option that is not itself chosen. Our model of the attraction effect shows that this is an adaptive behaviour resulting from using ordinal information to resolve uncertainty in determining which is the best alternative. It further makes a novel prediction that the attraction effect should be reduced in the loss domain. We conducted two experiments replicating and extending previous demonstrations of the attraction effect (Wedell, 1991; Soltani et al., 2012) and found the attraction effect in the gain domain but not in the loss domain. In choices between prospects, participants exhibited the attraction effect when prospects involved winning money, but not when prospects involved losing money. These results provide evidence that people are sensitive to the contextual information provided by ordinal relations and adaptively use this information to make better (higher expected value) decisions.

5.1 Introduction

The attraction effect Huber et al. (1982) shows how context can influence decision making. In one context a person chooses option $a$ over option $b$, but when a decoy alternative is added to the choice set, thereby changing the context, option $b$ is now preferred to option $a$. The attraction effect is typically regarded as ruling out rational value-maximising explanations of decision making. Our contention is that this behaviour actually results from an attempt to resolve uncertainty as to which alternative is the best in any given choice set and actually helps to maximise value. Our adaptive model of the attraction effect predicts that there should be a reduced effect in the loss domain. We test this prediction in replications and extensions of two experiments that have previously elicited the
The attraction effect is elicited in choice sets where two alternatives are described on two attributes (for instance car A and car B described in terms of price and quality) and where each alternative dominates the other on one of the attributes. It is possible to ‘attract’ choice toward either car A or car B by making that option a target. In order to make the option a target it is necessary to introduce a third alternative (the decoy) which is strictly dominated by the target on both attributes. The other alternative (the competitor) only dominates the decoy on one attribute. The decoy itself is not chosen but its introduction nevertheless biases choice away from the competitor and toward the target (see Figure 5.1).

This behaviour is problematic for value maximising models such as Luce’s (1959) choice axiom which are dependent on two axioms violated by the effect: Independence from irrelevant alternatives (IIA) and regularity. IIA stipulates that alternatives must be valued independently of one another, a consequence of this is that their valuation does not change when other alternatives are added to, or removed from, a choice set. Formally IIA means that:

\[
p(t\{t,c\}) \times p(c\{t,c\}) = p(t\{t,c,d\}) \times p(c\{t,c,d\})
\]

where \(p(c\{t,c,d\})\) is the probability of choosing the competitor given a choice set comprised of target, competitor and decoy.

Regularity simply requires that increasing the size of a choice set does not increase the probability of any of its original members being chosen: \(p(t\{t,c\}) \geq p(t\{t,c,d\})\). The attraction effect violates regularity because the target is more likely to be chosen when the choice set is expanded to include the decoy: \(p(t\{t,c\}) < p(t\{t,c,d\})\). The attraction effect violates IIA because the ratio of choices target over competitor changes when the decoy is added:

\[
\frac{p(t\{t,c\})}{p(c\{t,c\})} < \frac{p(t\{t,c,d\})}{p(c\{t,c,d\})}
\]

The fact that the attraction effect violates these axioms is widely cited when discussing value maximising decision models (Ariely & Wallsten, 1995; Heath & Chatterjee, 1991; Huber et al., 1982; Louie et al., 2013; Ratneshwar et al., 1987; Roe et al., 2001; Sen, 1998; Simonson, 1989; Tsetsos et al., 2010; Tversky & Simonson, 1993; Usher & McClelland, 2004). Consequently many models of the attraction effect focus instead on plausible mechanisms and processes that might produce the phenomenon (Roe et al., 2001; Usher & McClelland, 2001; Stewart et al., 2006; Bhatia, 2013). We have adopted an alternative approach which asks what the optimal solution is to a task in a given environment given the cognitive constraints that people must operate under. Our goal therefore is to explain the attraction as an adaptive response to the cognitive bounds we experience rather than an undesirable side effect of those bounds. We have developed such a model (Farmer et
al. submitted) which shows that the assumption of noise in expected value calculations is sufficient to model the attraction effect as an adaptive response to the inherent uncertainty in determining which alternative is best.

In order to provide strong evidence for this model we have sought to generate a novel prediction and test it against human behaviour, rather than simply reproduce existing phenomena. The model we have developed predicts that the attraction effect is adaptive in the gain domain, but much less so in the loss domain. If people exhibit the attraction effect because it is an adaptive behaviour rather than being the result of a particular heuristic or mechanism, then we would expect them not to exhibit a reduced, if any, attraction effect in the loss domain, since our analysis suggests that to do so could result in larger losses. If indeed people do not exhibit the attraction effect in the loss domain, this raises interesting questions around the predictive power of existing models of the effect. To take a simple example, Simonson’s (1989) contention that the effect occurs because it provides people with a justification for their choice, should apply equally well in the loss domain. In general, models that explain the effect by virtue of mechanisms that assess the distance between alternatives within an attribute should still predict the attraction effect in the loss domain. Our model does not make the same prediction in each domain because of a sensitivity to the underlying environmental properties of the attributes under consideration.

Our model was developed to explain the attraction effect in terms of expected value maximisation in choices among prospects. In order to make a prediction concerning the loss domain we simply, reversed the sign on the outcomes of the prospects and adjusted the ordinal relations accordingly. No changes are needed to the parametrisation of the model. The right hand panels in Figure 5.1 show a typical attraction effect choice set in the gain domain, while the left hand panels show a corresponding choice set in the loss domain. In the loss domain the decoy has a higher probability of losing a greater amount than the target, thus it is objectively worse on both attributes. If the attraction effect is present in choices between prospects in the loss domain then the target should be more likely to be chosen in the presence of the decoy since it dominates the decoy on both attributes - because it has a smaller probability of losing a smaller amount. In the following section we outline our model, and why it predicts a much reduced attraction effect in the loss domain. We then report two experiments designed to test this prediction.

5.2 A Model of the Attraction Effect as an Adaptive Response to Uncertainty in Expected Value Calculation

Our model starts with the assumption that cognitive processes are inherently noisy. Given this assumption we demonstrate that an expected value maximising decision maker can improve their decisions by taking into account the ordinal information implicit in attrac-
Figure 5.1: The attraction effect in gain and loss domains. In the right hand panels the target dominates the decoy because it has a larger probability of winning a larger amount. In the left hand panels the target dominates the decoy because it has a smaller probability of losing a smaller amount. In either of the domains an alternative A or B can be either a target or a competitor depending on the position of the decoy.
Combining two estimates of expected value, one derived from the metric information in the stimuli values, and the other inferred from the ordinal relations between those values\(^1\), naturally leads to the attraction effect. Importantly, it also leads to better (higher expected value) decisions than would be the case given an IIA consistent calculation of expected value. In the following sections we outline the model before showing that it makes different predictions in the loss and gain domains.

### 5.2.1 Uncertainty in Expected Value Calculation

Given noisy cognitive processes, an expected value maximiser will have some uncertainty as to which of two alternatives has higher expected value. When asked which of two prospects a person prefers, we assume that their expected value estimate of each alternative can be described by a normal distribution centred on the actual expected value and with a standard deviation that reflects their degree of certainty in their estimate. If these two distributions overlap then the decision maker has some uncertainty over which alternative is better. In this situation, any additional information that reduces uncertainty will result in higher expected value decisions. This is evident in Figure 5.2 where the right panel shows the proportion of correct choices (area under the curve to the right of the vertical dashed line) that result from the noisy estimates in the left panel. If the uncertainty in those estimates is reduced then the decision maker will make more correct choices, as in the second row of Figure 5.2.

### 5.2.2 Information Provided by Ordinal relations

One source of extra information comes from the ordinal relations among the alternatives in the choice set. In a two alternative attraction effect choice set containing prospects, the ordinal relations might be \(v(T) > v(C)\) and \(p(C) > p(T)\) where \(v(T)\) is the outcome value of the target prospect and \(p(C)\) is the outcome probability of the competitor. Given these ordinal relations, and no other information, there are two possible expected value orderings: \(E(T) > E(C)\) and \(E(C) > E(T)\) meaning that the ordinal relations imply an equal probability that either alternative has higher expected value than the other. When a decoy is included in the choice set the ordinal relations might be such that \(v(T) > v(D) > v(C)\) and \(p(C) > p(T) > p(D)\) where \(D\) is the added decoy alternative. With three alternatives in the choice set there are now six possible expected value orderings:

\[
E(T) > E(D) > E(C) \\
E(T) > E(C) > E(D)
\]

\(^1\)By ‘metric information’ we mean using the actual stimuli values, for instance 10% or £30. By ordinal information we mean the ranking of alternatives within an attribute, for instance if alternative \(a\) has a value of £10 and alternative \(b\) has a value of £20, then the ordinal information implies solely that \(b > a\).
Figure 5.2: Overlapping expected value estimates. In the left panel a decision maker has calculated an uncertain expected value for two prospects $A$ and $B$. The right panel shows the proportion of choices (to the right of the vertical line) that the decision maker selects the higher expected value alternative. The second row shows the same process for expected value estimates that have reduced uncertainty, consequently the higher expected value option is chosen more often.
A closer inspection of the ordinal relations reveals that not all of these ordering are possible since the target expected value must be greater than the decoy because it dominates the decoy on both attributes. The only possible orderings are

\[
E(C) > E(T) > E(D)
\]
\[
E(C) > E(D) > E(T)
\]
\[
E(D) > E(T) > E(C)
\]
\[
E(D) > E(C) > E(T)
\]

In all of these orderings the target has a greater expected value than the decoy. Crucially, however, in two out of the three orderings the target also has higher expected value than the competitor. We can therefore infer that, given just the ordinal information the target has a two thirds probability of having higher expected value than the competitor.

In attraction effect experiments it is not considered surprising that people do not choose the decoy. What is considered surprising is that the presence of the decoy leads people to choose the target over the competitor. The strict dominance of the target over the decoy allows the participant to infer with certainty that the target must be better than decoy, this reduces the choice to target or competitor. However, the above analysis shows that the exact same inference that eliminates the decoy also implies that the target is likely to be better than the competitor.

If something is known about the parent environment of the probabilities and values then it is possible to actually calculate an expected value estimate for each alternative using just the ordinal relations. If, for example, values are known to be normally distributed with a mean of £30, the ordinal relations we used above imply that the target will have a value higher than 30, the decoy will have the mean value of 30 and the competitor will have lower than mean value. The same principle applied to the probabilities (not necessarily normally distributed) results in the competitor having a larger than mean probability, and the target having the mean probability.

In Figure 5.3 we have conducted a simulation in which 1m values and probabilities have been generated. Three sets of 1m normally distributed values ($\mu = 20, \sigma = 8$) were randomly paired with three sets of 1m probabilities (Beta distributed, $\alpha = 2, \beta = 2$). The triplets of prospects that conformed to the attraction effect ordinal relations were then used to plot the expected values of the target, decoy and competitor (right hand panels). The left hand panels were generated in the same manner using pairs of prospects consistent
Figure 5.3: The attraction effect in the loss and gain domains. The distributions show the expected value of alternatives given just the ordinal relations between them. A is a target alternative and B is a competitor alternative. The left panels show the expected value of two prospects A and B when no decoy is present. The right panel shows that adding a decoy to the choice set changes the difference in expected value between A and B in the gain domain (bottom row) but not in the loss domain (top row).

with the attraction effect prior to a decoy being added. The bottom left panel shows that where there are only two alternatives these can be expected to have approximately equal expected values. The bottom right panel shows that when there are three alternatives in the attraction effect configuration, the target can be expected to have the largest expected value.

In summary, it is possible to use the ordinal relations in the attraction effect to infer a higher expected value for the target than the competitor. When something is known about the parent distributions of the attributes it is possible to put a value on the estimates of each alternative’s expected value. In the following section we demonstrate how combining this information with an expected value estimate based on the cardinal stimuli values leads to the attraction effect.

5.2.3 Combining Metric and Ordinal Information Leads to the Attraction Effect

In an attraction effect experiment involving choices between prospects, we assume that the decision maker derives a noisy expected value estimate of each alternative using the metric information provided by the stimuli. An expected value estimate of target and
competitor prospects in the attraction effect can be expected to lead to two overlapping distributions (not least because the stimuli are usually designed to have similar expected values). These overlapping distributions mean, as we have seen above, that the decision maker has uncertainty as to which alternative, target or competitor, is better. This uncertainty is represented by the spread of the difference distribution \( A - B \) (the dotted blue lines in Figure 5.4). The proportion of the difference distribution that is greater than zero reflects the probability that a decision maker would choose Alternative \( A \) over \( B \) given just the metric information and before the ordinal information has been taken into account.

The decision maker can reduce their uncertainty in their estimate by observing the ordinal relations in the choice set. As shown in the preceding section, this information can be used to infer the expected values of each alternative. The ordinal relations in a choice set imply a prior difference distribution between target and competitor (the dashed black line in Figure 5.4). A Bayesian integration of the metric and ordinal expected value estimates leads to a posterior estimate of the probability that either the target or competitor is better. Figure 5.4 shows that the posterior difference distribution (expected value of \( A \) minus \( B \)) changes according to whether alternative \( A \) is a target or a competitor. When \( A \) is the target (the left hand panels) the ordinal relations imply \( A \) is the better option, and therefore the prior weights the posterior in favour of \( A \). When \( A \) is the competitor (right hand panels) the prior weights the posterior in favour of \( B \). The proportion of the posterior distribution that lies to the right of zero in the top left panel of Figure 5.4 is the probability that the decision maker will choose option \( A \) when it is the target. The proportion of the posterior distribution that lies to the right of zero in the top right panel of Figure 5.4 is the probability that the decision maker will choose option \( A \) when it is the competitor. The change in the proportion of the posterior that is greater than zero between the left and right panels represents the size of the attraction effect - 27% in the example shown.

5.2.4 Reduced Attraction Effect in the Loss Domain

The exact same process described above, but in the loss domain produces a much reduced attraction effect. This happens because the ordinal relations imply the difference in expected value between target and competitor remains constant when a decoy is added to a choice set in which the target dominates on the probability attribute. This can be see in the top row of Figure 5.3 where the left hand panel shows that the expected value of an option \( A \) (dominant on the probability attribute) is roughly the same as the expected value of an option \( B \) (dominant on the value attribute). When a decoy is added to make \( A \) the target, the right hand panel shows that \( A \) and \( B \) still have roughly equivalent expected values. This means that, in the loss domain, when the ordinal relations reveal a target that dominates on the probability attribute, there is no implication that the target has higher or lower expected value than the competitor. This is reflected in the prior distribution in the
Figure 5.4: The attraction effect in the loss and gain domains. The left panels show the difference distributions between two alternatives when the target dominates on the value dimension. The right panels show the target dominant on the probability dimension. In the gain domain both decoy positions result in a prior that suggests the target has the highest expected value. The proportion of the posterior distribution to the right of zero represents the probability of choosing A. The difference in this area between the two decoy positions represents the preference reversal rate. In the top row of this figure the preference reversal rate would be 27%. In the loss domain only one decoy position suggests that the target is the higher expected value alternative. Consequently the preference reversal rate compared to the top panel, is much smaller. In this example the preference reversal rate would be half as much (13%). See Figure 5.1 for example configurations that correspond to each of the above panels.
bottom right panel of Figure 5.4 where the distribution has a mean of zero.

The bottom row of Figure 5.3 shows the expected value for a two-alternative versus a three-alternative choice set in the gain domain. The difference in expected value between the target and competitor alternatives (the dashed red and solid black lines in Figure 5.3) does change. Therefore the ordinal relations do imply that the target should be perceived as more attractive in the gain domain when it is dominant on the probability attribute. When a target is dominant on the value attribute the ordinal relations in both domains predict that it should be perceived has having higher expected value than the competitor. However, in the gain domain this information is consistent regardless of which attribute the target dominates on. In the loss domain this is not the case, since only one of the two possible decoy positions implies a more attractive target.

The bottom row in Figure 5.4 shows the attraction effect prediction for the loss domain given the same parameters as the gain domain on the top row. In the loss domain the attraction effect preference reversal rate is predicted to be just under half (13%) that in the gain domain. Our model therefore predicts that an adaptive use of the ordinal information should lead to an attraction effect in the gain domain but a much reduced one in the loss domain. These effects are robust to different environmental parameters, with the actual effect size being dependent on the coefficient of variance in the parent value distribution.

In the experiments that follow we confirm that people do indeed behave in this manner. We extend two existing studies showing the attraction effect in the gain domain (Wedell, 1991; Soltani et al., 2012) to include loss domain conditions.

5.3 Experiment 1

In Experiment 1 we adapted Wedell’s (1991) demonstration of the attraction effect in choices between prospects. The original study involved choices between prospects exclusively in the gain domain where participants chose between alternatives that all had an expected value of approximately $10. The alternatives were described in terms of probability of success and monetary value of success. These two attributes (value and probability) were traded off to produce attraction effect choice sets.

5.3.1 Methods

Participants

Sixty participants (42 male) were recruited using the Amazon mechanical turk platform. Participants had a mean age of 33 (SD = 11). Informed consent was collected and participants were paid $0.50. The experiment took approximately 10 minutes to complete.
Figure 5.5: Prospect values used in Wedell (1991) and in Experiment 1. Each prospect was paired with all others creating a total of 10 choice pairs.

Design

Within subjects, our independent variable of domain had two levels, loss and gain. Our dependent variable was preference reversal rate. This rate is a measure of the attraction effect based on that used by (Wedell, 1991). The rate an alternative is chosen when it is the competitor is subtracted from the rate the same option is chosen when it is the target. The attraction effect is present when an alternative is chosen more often when it is in the target configuration than in the competitor configuration. The same alternative is presented several times in each configuration, hence it was possible to calculate a rate for each.

In each of the gain and loss domains we presented participants with 10 pairs of prospects, one high probability (safe) prospect with a value ranging from $12 to $25, and one low probability prospect (risky) with values ranging from $15 to $33 (see Figure 5.5). All prospect pairs had an expected value of approximately $10. In the loss domain the gambles were identical except that the value amounts were negative.

In both domains we used range decoys which was one Wedell’s (1991) between subjects conditions. Range decoys have previously been shown to elicit the largest attraction effect size (Trueblood et al., 2013) and consequently provided the strictest test of our hypothesis that the effect would be reduced in the loss domain. In the gain domain this meant that the safe prospect decoys had a $2 smaller value, and the risky prospects had a 5% smaller probability. In the loss domain, the safe prospect decoys had a 5% greater probability and the risky prospect decoy had a $2 greater value.
In each domain the 10 safe and risky prospect pairs were presented twice. Once with the safe prospect as the target and once with the risky prospect as the target. The decoy position defined whether a prospect was the target or competitor.

**Stimuli**

In the gain domain, three prospects were presented simultaneously (target, competitor and decoy) in sentence form: ‘83% probability of winning $12’. Each choice set was prefixed with the question: ‘Which option do you prefer?’. In the loss domain the sentences took the form ‘83% probability of losing $12’. The word ‘losing’ was coloured red to indicate that domain had changed.

**Procedure**

Participants answered 40 questions in total. Within each domain the order that the choice sets appeared in was randomised. The loss and gain domains were presented as separate blocks. Participants experienced each choice set as a separate web page. On each page they indicated their choice by pressing the appropriate radio button. They then pressed a continue button to advance to the next page. A page counter kept participants informed of their progress through the experiment.

**5.3.2 Results**

Figure 5.6 shows the size of the attraction effect in each of the domains. The preference reversal rate was 10% in the gain domain and not significantly different from zero in the loss domain. The choice proportions recorded were arcsin transformed and revealed a significant attraction effect in the gain domain using a two-tailed one-sample $t$ test, $t(53) = 3.92, p < .001, d = 0.53$ but not in the loss domain, $t(53) = -0.49, p = .63$. A paired samples two-tailed $t$ test revealed that the preference reversal rate in the gain domain was also significantly different from the rate in the loss domain, $t(20) = 3.03, p = .004, d = 1.21$

**5.4 Experiment 2**

In Experiment 2 we extended a more recent demonstration of the attraction effect in choices between prospects (Soltani et al., 2012). Whilst the Wedell (1991) experiment equated the expected value of the alternatives on offer, it should be noted the risk preferences are likely to mean that participants did not in fact find these alternative subjectively equivalent. The Soltani et al. (2012) study resolved this issue by having a preliminary
<table>
<thead>
<tr>
<th>Condition</th>
<th>Preference reversal rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain</td>
<td>0.10</td>
</tr>
<tr>
<td>Loss</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Figure 5.6: Results of Experiment 1. The attraction effect was present in the gain domain but not in the loss domain. Error bars are standard error.

A session in which each participant’s point of subjective equality was found for a fixed high probability prospect.

### 5.4.1 Methods

#### Participants

Twenty-eight undergraduate students (3 male) from the University of Manchester volunteered to take part. Participants had a mean age of 21 ($SD = 1.8$). This sample size was chosen to match that of the original study (Soltani et al., 2012). Informed consent was collected and participants received course credit for taking part. The experiment took approximately 45 minutes to complete.

#### Materials

The experiment stimuli were presented on a 17 inch display at a resolution of 1280 x 1024 pixels. The experiment was programmed in Python and run on Windows 7. Participants responded using the numeric keypad of a standard Windows keyboard.

#### Design

As per Experiment 1, we used a within subjects design, testing an independent variable of domain (loss and gain). Our dependent variable was preference reversal rate. The dependent variable was measured in exactly the same way as in Experiment 1.
In each of the gain and loss domains we presented participants with two prospects, one high probability (safe) prospect with a value of either £20 or -£20, and one low probability prospect (risky) that had a 30% probability of winning an amount that each participant found subjectively equivalent to the safe prospect.

We determined a point of subjective equality for each participant using a two-alternative forced-choice paradigm. We paired a fixed safe prospect with a risky prospect and systematically varied the value of the risky prospect. This method is adapted from (Soltani et al., 2012). The value varied from £20 to £120 in ten steps of £10. Each step was repeated eight times. This allowed us to fit a logistic curve to the data and recover the indifference point such that the participant would be equally likely to choose either prospect. This approach is commonly used in studies of perception to recover perceptual points of subjective equality (PSE) of stimuli. The recovered PSE was then used for the risky prospect in the main study. This process was carried out for both the gain and loss domains. In each domain the safe and risky prospects were presented nine times with the safe prospect as the target and nine times with the risky prospect as the target. The decoy position defined whether a prospect was the target or competitor.

Stimuli

Participants were presented with prospects on a computer monitor. The screen was divided vertically into three panels. Each panel showed a probability expressed as a percentage and a value prefixed by a £ sign. In the gain domain the values were coloured green, whereas in the loss domain the values were coloured red and prefixed by a minus sign. During the evaluation phase of each trial the word Evaluate appeared in white at the top of the screen. During the choice phase of each trial the word ‘Choose’ appeared in yellow.

Procedure

The experiment started with a session to determine each participant’s PSE. Participants were presented with the safe prospect and a risky prospect with a random value until all the risky prospects had been presented eight times. Participants could observe the prospects for as long as they liked. When they were ready to indicate their choice they pressed the space bar and then had two seconds to respond by pressing the appropriate number on the numeric keypad (1 for left panel, 2 for middle and 3 for right panel). As per Soltani et al. (2012), a logistic curve was then fitted to the data in order to calculate the PSE.

In the choice phase participants were instructed that after evaluating the prospects they would press the space bar and one of the prospects would be removed at random. They then had two seconds to choose from the remaining prospects. In the majority of trials, the
decoy was removed and participants chose between the target and competitor. However, in some dummy trials the decoy remained. These trials were discarded but forced the participants to consider all three alternatives on each and every trial since they could not know which prospect would become unavailable.

In both phases of the experiment the risky and safe prospects were slightly jittered such that participants were not presented with same choice twice. For each prospect the value could vary by £1 and the probability by 1%. There were 36 trials of interest in the choice phase (2 domains x 2 decoy position x 9 repetitions). These were supplemented with a further 18 dummy trials. The trials were presented in a random order but were blocked by domain.

5.4.2 Results

Seven participants were excluded from all analyses presented here since their indifference points either exceeded the range of values we offered, or because their indifference value did not leave sufficient space for a decoy placement that conformed with the attraction effect ordinal relations.

In both domains it was possible to fit logistic regressions to the participants’ choice data. Figure 5.7 shows an example of choice proportion data in each domain. These data indicate that participants understood the task and responded predictably in both domains preferring smaller losses and larger gains. The mean PSE was -£59 in the loss domain and £70 in the gain domain.

Figure 5.8 shows that there was a significant attraction effect in the gain domain with a preference reversal rate of around 11%. A two-tailed one-sample t test was conducted on the arcsin transform of the preference reversal rate to determine whether it was significantly different from zero, \(t(20) = 2.19, p = .04, d = 0.48\). The preference reversal rate was not significantly different from zero in the loss domain, \(p = .76\). The preference reversal rate in the gain domain was also significantly different from the rate in the loss domain as determined by a paired samples two-tailed t test, \(t(20) = 2.71, p = .01, d = 0.55\).

5.5 Discussion

We aimed to test a novel prediction of our adaptive model of the attraction effect. The model predicts that the attraction effect should be present in the gain domain but much reduced in the loss domain. In Experiment 1 we replicated the findings of Wedell (1991) showing the attraction effect in choices between prospects in the gain domain. The same subjects did not exhibit the attraction effect when presented with the same prospects but framed as losses.

In Experiment 2 we replicated and extended Soltani et al. (2012) showing that the
Figure 5.7: Example choice data for loss and gain domains. The data shown are from one participant and a logistic regression has been fitted to their data in order to retrieve a risky (30% probability) gamble value that they determined to be equally attractive to a 70% probability of £20.

Figure 5.8: Preference reversal rate in the loss and gain domains. Error bars are standard error.
attraction effect was present in choices between prospects in the gain domain. The same subjects did not exhibit the effect in the loss domain, despite in both domains using pairs of prospects that participants found to be equally attractive. Although these experiments were designed as tests of our model, it is an interesting (and to our knowledge, novel) finding that the attraction effect is not present in choices between prospects in the loss domain. We discuss the support that this finding provides for our model below.

These data provide further evidence for our model. The model was developed simply to account for the standard attraction effect in the gain domain by integrating expected value estimates derived from the metric and ordinal stimuli values. We were entirely unaware that this would result in a different prediction for the loss domain. The fact that this is a novel prediction of the model, and is borne out empirically, suggests that the model has some explanatory power.

In the loss domain, when a target alternative is dominant on the probability dimension, the ordinal relations imply that the difference in expected value between target and competitor remains constant in the presence or absence of a decoy. By contrast in the gain domain, the presence of the decoy always means that the ordinal relations imply the target has the higher expected value than the competitor. According to the model therefore, it is adaptive to exhibit the attraction effect in the gain domain but may not be in the loss domain. That this is the pattern people exhibit, suggests that the attraction effect can be understood as an attempt to make value maximising choices.

Our results show a preference reversal rate of zero in the loss domain, whereas the model predicts that the effect should be reduced relative to the gain domain, but not eliminated. If the model is set to the environmental parameters of the Wedell (1991) stimuli and the noise in expected value calculation parameter is free to vary, then for an 11% preference reversal rate in the gain domain, a 6% rate is predicted in the loss domain. There are several plausible reasons for this discrepancy. First and foremost, it should be noted that the model has just one free parameter (the agent’s noise in calculating expected value), and has just three parameters in total. As noted, the environmental parameters (mean and variance of prospect values) were taken from the Wedell (1991) stimuli. There are several further plausible parameters which would reduce reliance on the environmental priors and thereby reduce the effect further in the loss domain. One such parameter would be to introduce noise on the perception of the ordinal relations such that there was some possibility the decision maker would make a mistake in perceiving the ordinal relations. However, there is value in predicting the qualitative effect we have found with as few parameters as possible. The model seeks to explain why the behaviour is adaptive rather than account for the procedure by which people process the stimuli they are presented with. People may not achieve the optimum performance but nonetheless be broadly shaped in their decision making by the optimum course of action.

If our model is an accurate representation of the underlying reason that people exhibit
the attraction effect then several further possible streams of research might be productive. The model assumes people have some degree of sensitivity to the background, or parent distributions of attributes in a given decision environment. In the case of the experiments reported here this may well reflect the kinds of values and probabilities that people experience in psychology laboratory and online experiments. One plausible test of this assumption would be either to train people in an environment in which we have manipulated the distributions, or to measure distributions of attributes in some known ‘real-world’ environment.

One intriguing issue raised by our results is that even a purely expected value maximising agent may behave differently in the loss domain to the gain domain given otherwise identical statistical properties. In a more formal implementation of this model (Howes et al., 2014) we show that an expected value maximising process subject to noise will exhibit risk preferences without any need to specify a risk preference parameter.

5.5.1 Conclusion

These findings highlight the potential explanatory power of trying to understand behaviour as optimal within bounds. Our approach starts with the assumption that human cognition operates under constraints, but then asks what the optimal behaviour is given the task environment and the constrained abilities of the agent (Howes et al., 2009; Lewis et al., 2014). The potential benefit of this method is that we can establish a causal explanation for decision making phenomena like the attraction effect. One way to test the validity of this approach is to generate novel predictions and test whether these have empirical support.

We have provided evidence that the attraction effect does not occur in the loss domain, and more importantly supported our claim that the effect is an adaptive response to noisy cognitive processes.
Chapter 6

Discussion

6.1 Summary of Findings

Objective 1 was to identify the conditions under which rational models are violated and what the consequences of these violations are. Chapters 2 and 3 have provided empirical evidence for both the consequences and limits of the violations of IIA and regularity implicit in the attraction effect. Chapter 2 showed that the attraction effect can be elicited in a rapid pointing task. Participants learned their own motor noise and were then asked to choose between different aim points in an attraction effect configuration. In a control condition participants were also asked to choose between prospects presented in sentence form. In both conditions participants chose an alternative more frequently when it was in a target position than when it was in a competitor position. Although motor tasks of the type used in Chapter 2 often reveal near optimal performance, that is because they are measured in terms of efficiency. Were the data from attraction effect experiments also measured in this way it would reveal that the supposed irrational behaviour has little consequence for the decision maker.

In Chapter 3 participants were asked to choose between prospects presented using a variety of paradigms. In all of the paradigms the same probabilities and values were used to construct choice pairs that varied in expected value difference from zero to 300%. A high level adaptive model predicted that the attraction effect should decline with increases in expected value difference, and that the effect should decline more rapidly in the paradigms where people found it easier to perceive the difference in expected value. Our findings confirmed these predictions. The reduction in preference reversal rate was found across different presentation modes which included judging the area of rectangles as well as stating a preference between explicitly described prospects. The preference reversal rate declined more rapidly as a function of expected value difference for the rectangle condition than it did for the descriptive condition. The high level model presented in this chapter predicts this pattern since it is assumed that judging the difference in area
between rectangles is subject to less variability than in calculating the difference in expected value between prospects described in sentence form. This sensitivity to expected value difference reveals that the violation of rational models is limited to cases where alternatives already have similar value. In subsequent sections we will discuss why these results, although predictable, have important consequences for our interpretation of the attraction effect.

Objective 2 concerns developing a computationally rational model of the attraction effect. Chapters 3 and 4 provide a high-level and more formal version of such a model respectively. The model in Chapter 3 provides a link between objectives 1 and 2 in showing that the attraction effect is beneficial, whilst simultaneously establishing the conditions under which the effect can be elicited. The model shows that the attraction effect is beneficial, whilst simultaneously predicting the results of the empirical test used to establish under what conditions the effect could be elicited.

The more formal model in Chapter 4 was successfully fitted to the Wedell (1991) attraction effect data. This model shows that a Bayesian integration of expected value estimates provided by both the ordinal information and a noisy expected value calculation will naturally lead to the attraction effect. Furthermore this model will also produce the compromise and similarity effects. A further insight provided by the model was that the risk preference evident in Wedell’s data emerged naturally from our model without any need for a risk preference parameter. This then raises the possibility that the computationally rational model of decision making presented to account for the attraction effect has the potential to explain a range of apparent biases in the literature.

Chapter 5 addressed the final objective of testing novel predictions made by our model against human data. The model from Chapter 3 somewhat counter-intuitively predicts that the attraction effect should be reduced in the loss domain. To test this prediction we replicated two existing studies using prospects in the gain domain, (Wedell, 1991; Soltani et al., 2012) and added loss conditions. Analysis of these data revealed that participants perceived the loss conditions in the manner intended since they chose the risky loss alternative progressively less often as the value of the of loss increased. In both experiments we succeeded in replicating the gain domain attraction effect, but did not find any attraction effect in the loss domain, consistent with the model predictions.

6.2 Synthesis

This section returns to the three objectives identified in Chapter 1 and addresses in detail how the above findings both provide answers and suggest new avenues for research. For each objective an interpretation of the results and the limitations of the work are identified. These two aspects are informed by previous and current work by other authors in this area. Finally, further research is suggested.
6.2.1 Objective 1: Establish the Consequences of Violating Rational Axioms

The core aim in Objective 1 was to understand the conditions under which people will violate the axioms of value-maximising models. It also sought to understand what the consequences of violating these axioms might be for a decision maker. In essence this was about understanding when someone is likely to exhibit the attraction effect and what difference it actually makes to them in terms of the outcomes they subsequently experience. To state that decision making violates the axioms of rational models might conjure up an image of a somewhat chaotic process in which the decision maker might unpredictably choose the worst alternative on offer. In reality, the conclusion from pursuing Objective 1 is that the attraction effect occurs under certain limited circumstances in which the possible negative consequences are negligible. In fact by a different metric of rationality one might not even notice anything remiss in the attraction effect whatsoever. These arguments are developed below.

The decision making literature on rapid pointing tasks has suggested that people can make near optimal decisions in these paradigms (Trommershauer et al., 2008). This raises the intriguing question of whether people would exhibit the attraction effect in these types of task. If people behave rationally when prospects are presented as pointing tasks but do not behave rationally when prospects are presented in an attraction effect configuration, it was plausible that pointing tasks would eliminate the attraction effect, or conversely, that the attraction effect configuration would reduce the near optimality of pointing paradigms. The data revealed that people do exhibit the attraction effect in pointing tasks. However it is problematic to regard this as evidence that the attraction effect reduces the optimality of decisions in pointing paradigms. To see this it is necessary to examine the different ways in which rationality is measured in different decision paradigms.

In the rapid pointing tasks now widely used (Trommershauer et al., 2008; Wu et al., 2009; Jarvstad et al., 2013) the measure of performance has typically been the expected reward that would be achieved by participants’ mean aim points. Trommershauer et al. (2008) have shown that participants typically aim for a point that achieves near to the optimum payoff. The rationality of their decision is measured in terms of efficiency, i.e., they achieve a high percentage of the maximum possible score they could achieve. However, Chapter 2 reveals that despite this efficiency the same paradigm can be used to elicit the attraction effect. There is no contradiction here, the axiomatic violation of value maximising models implicit in the attraction effect has a negligible impact on the actual efficiency of a decision (later it will argued that it actually increases it).

Perhaps the best way to illustrate why it has negligible impact is to consider the results of Chapter 3. Here it can be seen that as the potential cost to efficiency of exhibiting the
attraction effect increases, so people decrease the extent to which they exhibit the effect. In fact, the attraction effect peaks precisely when there is no difference at all between the efficiency of the alternatives on offer (because there is zero difference in expected value).

That the size of the attraction effect is driven by the difference in expected value between two prospects suggests that the common perception that it violates value maximising models should be given a more subtle reinterpretation. In fact the attraction effect is sensitive to the extent that value maximisation yields a clear choice. This is the primary motivation behind Chapter 3, to show that value maximisation is only violated in choice sets where value maximisation leaves ambiguity as to the best alternative.

The model presented there is not unique in suggesting that making the expected value of choices discernible will lead to a reduced attraction effect. This assumption is implicit in the experimental design of all attraction effect experiments, and the model also does not make a prediction with respect to expected value differences which would not also be made by all other models. However, we started with a different end goal in mind, not to provide an alternative process model, but to provide an explanation of why the attraction effect is an adaptive behaviour. Clearly, if the attraction effect were to persist in the face of large differences in expected value then it could not be regarded as adaptive. We simply provide empirical evidence that it does not persist. This evidence underpins the assertion that value-maximising models are violated when value maximisation is already guaranteed.

The impact of the effect is that it supposedly reveals that people do not value-maximise, however it only reveals this in choice sets where value maximisation is virtually guaranteed because the alternatives are designed to be equally valuable. This point has not been lost on economists interested in the practical relevance of such psychological phenomena:

[we should] conduct experiments which show not merely that individuals are sometimes systematically non-rational, but which would start to give us some feeling for how important this phenomenon might be. For example, I don’t think that any of the preference reversal experiments have yet given information that would allow us to graph the frequency of preference reversals as a function of the difference in expected value (in, say, percentage terms) between the two gambles with which subjects are presented. While this information might not dramatically effect how investigators with very different theoretical dispositions evaluated the data, it would help address the issue of whether we are seeing a phenomenon likely to play an important role in natural environments. Undoubtedly there are more informative experiments to be done than this one; my point is only that investigators who are aiming to assess the size of the disparity between utility theory and observed behavior, rather than merely to show that there is one, might better address the usefulness of the utility maximization approximation. (Roth, 1996, p. 202)

Roth’s perspective bears a strong resemblance to a rational analysis perspective on
decision making phenomena. This is likely because both are concerned with explanations in terms of the pressure bought to bear by the environment external to the individual. From a rational analysis (Anderson, 1990) perspective we might expect decision making to conform to a measure of rationality that maximises gain rather than one that exhibits internal logical consistency. In fact Anderson (1990) argued along strikingly similar lines in his discussion of the implications of violating transitivity:

A question that is rarely asked is whether there is really a cost associated with the purported irrationality. If a person prefers A to B, B to C, and C to A, but there are no differences between A, B and C in their adaptive value, then the intransitivity does not violate the adaptive principle of rationality. (Anderson, 1990, p. 32)

Chater and Oaksford (2000) argue that apparently irrational behaviour may be simple to explain if a different normative solution is applied to the problem. The findings in Chapters 2 and 3 lend some support to the wrong normative standard being applied in the case of the attraction effect. Chapter 2 shows the axiomatic violation of the attraction effect in a pointing task, while Chapter 3 shows a sensitivity to efficiency in a choice paradigm traditionally used to elicit the attraction effect. In both experiments performance could be labelled as rational or irrational depending on which measure of performance is used.

The primary contribution of Chapters 2 and 3 is to highlight the importance of applying the correct standard against which to judge decision making. The attraction effect can be elicited in a task typically regarded as demonstrative of optimal human decision making. Also, in traditional attraction effect experiments people can be made to choose in accordance with value maximising axioms simply by providing them with choice sets where value maximisation is actually possible. This suggests that an efficiency metric is a better normative standard against which to judge decision making.

The presence of the attraction effect in the results of Chapter 3 provide no evidence of a violation of the adaptive principle of rationality. However, the problem with this conclusion is that it does not really explain why the attraction effect is exhibited in the choice problems where there is little to be lost. This is the domain of Objective 2 which will be discussed in the next section (6.2.2) and will be the core contribution of this thesis. It is argued that exhibiting the attraction effect increases efficiency given the cognitive constraint of noise. Before moving on to that objective it is worth noting that a normative standard of efficiency sees no contradiction in the attraction effect if it can be shown to improve the expected value of decisions made. This contrasts with the axiomatic definition, which will remain violated whether or not efficiency is increased by the attraction effect.

A limitation of the rapid pointing task in Chapter 2 was that our participants did not have to make any movements in the actual decision phase, instead they were given a
hypothesized choice and asked which of the targets they would aim for. It was assumed that they represented the width of the bars in terms of their probability of hitting them given the motor noise they had been trained in. Our primary reason for using this method was that it had been employed successfully in previous experiments (Wu et al., 2011; Jarvstad et al., 2013). In addition, Wu et al. (2011) have provided neuroimaging evidence that participants do appear to treat this task as a hypothetical motor decision. Specifically, in an fMRI study they showed that activity in the medial Pre-Frontal Cortex correlated with probability weighting in motor decision tasks but not with bar size in a task using identical stimuli but judgements of bar width.

In Chapter 3 it is assumed that judging the difference in areas of rectangles is subject to less variance than a mathematically equivalent judgement of prospect expected values presented in sentence form. This seems like an uncontroversial assumption, but is an assumption nonetheless. An improvement to this design, and one which would allow us to test the model more rigorously, would be to measure the precision with which people judge differences in area as opposed to differences in expected value. One way to achieve this would be to follow a similar method to that used in Experiment 2 of Chapter 5. Using a 2AFC design, participants would be asked to judge either the larger of two areas or better of two lotteries. From this data it would be possible to recover a psychometric function for each type of discrimination and from the slope of that function estimate the precision of the estimate. This would also allow us to make predictions about individual differences between participants within a paradigm. For instance, a participant who had a steeper psychometric function as a result of varying the expected value of prospects would be expected to show a more rapid decline in their attraction effect rate as a function of increasing expected value differences.

Future research should address the above issues raised by Roth and Anderson concerning the consequences that these apparently irrational decisions actually have. Assessing decision making phenomena in terms of efficiency would be a good starting point. How much difference in expected value between alternatives would be tolerated before many of the phenomena such as the Allais paradox ceased to hold? That value maximising models break down in choices where it is not evident which alternative is best suggests that as a strategy they leave ambiguity with regard to the correct course of action (perhaps because of noisy cognitive processes) and so any divergence from rational models is likely to be a sensible attempt to deal with that ambiguity. A reading of the decision making literature, however, leads to the interpretation that people are in no way capable of value-maximisation.

One way to explore a different metric would be to examine the signal to noise ratio, in terms of efficiency, that a value-maximising strategy achieves in different choice problems. This could lead to predictions concerning when a person is likely to depart from value maximising strategies. In such conditions, we might then ask what is the solution,
given the unresolved uncertainty, to obtaining extra value from the decision problem. Optimal analyses from this starting point might reveal, as we have done for the attraction effect, that contextual information such as choice set composition can be integrated to reduce uncertainty. This would stand in contrast to the all or nothing categorisation of people as either value-maximising or not, that measurement against axioms in zero expected value difference choice sets has so far provided.

6.2.2 Objective 2: A Computationally Rational Model

Objective 2 was to develop a computationally rational analysis of the attraction effect. This method seeks to establish a causal link between the computational capabilities of an agent and their behaviour by understanding what the optimal behaviour is given those computational abilities. In the models developed in Chapters 3 and 4 we simply assume that people have noise in calculating expected value. This allows us to show that the ordinal relations among the alternatives in an attraction effect choice set provide information that can increase a person’s accuracy in calculating expected value. A consequence of using this information is that people will exhibit the attraction effect as well as the compromise and similarity effects. Exhibiting the attraction effect is an adaptive behaviour once the computational constraints of the decision maker are factored in to the decision problem.

The key insight this method has provided is that the effect is adaptive. This suggests that people exhibit the attraction effect because it helps them make better decisions. This allows us to start to answer why, rather than how, people exhibit the attraction effect. The literature on the attraction effect to date has focussed on explanations that propose mechanisms, algorithms and heuristics that would produce the effect (Huber et al., 1982; Simonson, 1989; Tversky & Simonson, 1993; Roe et al., 2001; Usher & McClelland, 2004; Stewart, 2009; Bhatia, 2013). These explanations are not refuted by the account presented here, rather, this account provides a reason why those mechanisms and processes might be advantageous. Perceiving the attraction effect as adaptive is potentially more insightful than developing models of mechanisms that can reproduce it. Whilst the models such as MDFT and AAM can reproduce an impressive number of phenomena they are essentially descriptive. The phenomena are explained in terms of the mechanism, but the model of the mechanism is derived with a view to reproducing the phenomena. By contrast, our account proposes that people exhibit the attraction effect because they have no choice given their cognitive constraint of noise and a goal of maximising expected value. It is the ‘no choice’ part of the previous sentence that reveals the benefit of the computationally rational approach. Because behaviour is explained in terms of achieving a goal we can identify plausible constraints that would result in the behaviour we see given that goal.
Noguchi and Stewart (2014) show that eye-tracking data can provide evidence in favour of some contextual preference reversal models over others. For instance, their data reveal that people seem to make comparisons between pairs of alternatives within an attribute. These data provide evidence in favour of models such as decision by sampling (Stewart, 2009) but are problematic for models such as AAM (Bhatia, 2013). It is difficult to determine the extent that these data can challenge the model presented here. First and foremost the model is not a process model and therefore does not specify an order of operations, or indeed how information should be encoded. Certainly the Noguchi and Stewart (2014) findings do not rule out the calculation of expected value which could be accurately derived by a number of different algorithms, or approximated by infinitely many. It is perhaps not useful to ask whether eye tracking data can determine whether or not people make expected value calculations. The place of expected value in the model is to provide an external pressure which shapes behaviour.

That these data do not refute our model raises interesting questions about precisely how the model could be refuted. It is however, not the case that abstracting away from mechanism makes rational models unfalsifiable, although this is a criticism that has been levelled at Bayesian models in particular (Bowers & Davis, 2012; Jones & Love, 2011, although see Hahn, 2014 for a rebuttal). In fact, the model could be refuted in a number of ways. For instance, it makes specific predictions about the underlying statistical environments which determine the rational behaviour given the ordinal relations. An empirical test could be conducted by training people in an environment, showing that they are aware of its properties, and then contrasting behaviour with that predicted by the integration of ordinal information with expected value calculation. The experiments in Chapter 5 also provide a test of our model since, had people exhibited the attraction effect to the same extent in the loss domain, the empirical finding would have contradicted the model.

One unexpected insight that the model revealed was that the risk preferences evident in the Weddell (1991) data naturally emerge from our model which assumes only that people are maximising expected value. The calculation of expected value does not normally permit any degree of risk aversion unlike expected utility which has a parameter to achieve precisely this. The reason the model presented here exhibited risk aversion was due to the interaction of maximising expected value and the underlying statistics of the choice environment. This raises the prospect that risk aversion can be understood as a rational behaviour (expected value maximisation) that naturally emerges in certain environments. While further work needs to be done to understand the extent that this finding might be generalised to other risky choice phenomena, there is now a growing literature which can explain apparently non-rational decision making with reference to the broader statistical properties of the environment the agent is in (e.g., Hahn & Warren, 2009). Particularly relevant to this research is recent work by Fawcett et al. (2014); Trimmer (2013) who show that many apparently irrational decision making phenomena in animals can be explained
as an adaptive response to the statistical structures of the complex environments they inhabit.

Since the model relies on an assumption of sensitivity to the underlying environmental distributions, this suggests an obvious area for further work detailing how the nature of parent distributions affect the predicted rationality of exhibiting the attraction effect. An obvious test of the model which suggests further work is to manipulate the underlying environmental distributions from which the alternatives in a choice set are sampled. This could be done in several ways. One alternative is to design a novel task in which we specify the statistical properties of the environment, and train people in that environment. This would allow a between subjects manipulation in which we test whether the changes to the statistical environment produce predicted changes in preference reversal rate.

The computational rationality approach argues that people will adjust to the optimal strategy given a particular payoff function and feedback on the consequences of their decisions. While the above potential experiments suggest that feedback would be derived in the long term from interacting with an environment, separate predictions could be obtained by manipulating the nature of the payoff function and feedback in an environment controlled by the experimenter. In previous work for instance, this approach has been used to show that people adapt to the optimal point to interleave tasks (Farmer et al., 2011; Janssen et al., 2011).

An alternative approach to specifying an environment, would be to measure the properties of a particular real life environment. In a property market the attributes of distance from town centre and price could be used to construct choice sets and then test for contextual preference reversal in that market as compared to another market with different attribute distributions. The underlying statistical properties of each environment would predict differing preference reversal rates that could subsequently be tested empirically.

A further potential extension of this work is to consider how the cognitive constraint of noisy valuation and sensitivity to statistical properties of the environment might be developed in to a general theory of decision making. Under this approach it might be possible to predict decision making phenomena resulting from the optimal integration of noisy valuations of stimuli, together with contextual information which reveals additional information given the assumption of sensitivity to environmental statistical properties. A unified theory of decision making might emerge in which people carry out rational value-maximisation informed by the statistical structure of the environment they are in.

Other aspects of the model presented here present some potential difficulties in terms of known interactions with the attraction effect. Recent work by Pettibone (2012) purports to show that as the evaluation time increases the attraction effect also increases. Consequently, the longer a person spends evaluating a choice set the more likely they are to exhibit the effect. Pettibone shows that the ratio of target to competitor choices increases as the evaluation time increases from two seconds to eight seconds. These re-
sults are predicted by MDFT (Roe et al., 2001) and LCA (Usher & McClelland, 2004) which also show that the target should be more likely to be selected as time passes. A potential criticism of the rational model presented here is that it seems plausible that the longer a person spent evaluating the alternatives the more accurate their expected value estimate might be and consequently the smaller the model predicts the attraction effect should be. However, this criticism is potentially unwarranted. Viewed from a different perspective the Pettibone data reveal that from a ceiling attraction effect in the 8 second condition it is possible to reduce the size of the effect by imposing increasingly severe time constraints. However, a plausible reason for this is simply that the time constraint adds random noise to the participants’ behaviour. The results reveal as much since in the two second condition participants are choosing the decoy 20% of time. This implies that given a very severe time constraint people will simply choose randomly, i.e., one would expect the target, decoy and competitor each to be chosen a third of the time.

Furthermore the rational model presented in this thesis is not a process model, it is a model of what it is rational to do once information has been assimilated. If severe time constraints were modelled they would result in increased noise on the perception of the ordinal relations (as the Pettibone data suggest) and consequently there would be a trivial prediction that the attraction effect would decrease as choices tended toward random. Indeed Figure 4.6 shows the effect of increasing noise on the extent of the attraction effect. It is in fact interesting to evaluate MDFT and LCA with respect to time course. These models argue that predicting the time course of a decision is a significant benefit of their approach. However, it seems very difficult to disprove the prediction that decision phenomena will take time to manifest themselves. Static models do not assume that people make instantaneous decisions, and are not invalidated by the fact that it takes time for a person to observe, encode and process stimuli.

One critical aspect of the model presented in Chapter 4 is that it assumes the two attributes used to describe the stimuli in an attraction effect experiment are independent of one another. That is to say, they are not correlated. One might reasonably expect that many attributes used in attraction effect experiments are in fact negatively correlated, for instance the price and quality of most products, or acceleration and fuel-efficiency in cars. On the other hand this is not necessarily the case for all stimuli that have been used, such as restaurant food quality and distance. It is interesting to note that the three product categories in Huber et al. (1982) that elicited the largest attraction effect were described on attributes that might reasonably be perceived as not being correlated. These were fuel-efficiency and ride quality in cars, food quality and restaurant distance, and TV picture distortion and set reliability. Initial simulations in an environment with correlated attributes suggest that the stronger the (negative) correlation between the attributes, the larger the expected reduction in the attraction effect. The extent that the underlying statistics of the environment would affect decision making is a prediction of the model that is
empirically testable.

A related assumption is that people perceive the offered alternatives as random samples from their parent distributions. In other words it is assumed that people perceive the offered alternatives as being representative, and randomly selected from the market or environment in which they are found. It seems likely that this is actually implicit in most attraction effect experimental designs. The alternative perspective would be that the participant assumes the experimenter is deliberately placing the decoy option in the choice set. The stimuli question sheets in both Huber et al. (1982) and Wedell (1991) ask participants to imagine they are faced with the following choices. This imagination process, very plausibly includes the assumption that the products on offer are representative of a parent marketplace. This assumption could also plausibly hold even if the participant is aware of the experimental design. Such an awareness would not prevent them from choosing how they would behave were these genuine offerings encountered in the real world. In fact the alternative point of view, that people approach the task as a purely psychophysical exercise, well aware that the choice set has been constructed by the experimenter to conform to the attraction effect ordinal relations, seems the most unlikely scenario. If no participant has ever perceived the alternatives in an attraction effect experiment as being sampled from some real or imagined environment, then arguably, the data are meaningless since it is a demonstration of what people do when cognisant of an experimenters intentions.

The model is based on the integration of the ordinal and metric information presented in the stimuli. It is questionable to what extent these can be considered to be two separate sources of information. The metric information about the value of each alternative on each attribute is precisely the source of the ordinal information. In other words, without the metric information there is no ordinal information. Simulations reveal that combining estimates based on separate perception of the metric and ordinal sources do indeed lead to a higher rate of correct choices.

In teasing apart this issue, it is instructive to look at the normative solution to the problem of choosing between gambles. In a choice between gambles the unconstrained rational decision maker simply chooses the gamble with the highest expected value. The cognitive constraint assumption in the model is that people are subject to noise in this process leaving uncertainty as to which of the alternatives genuinely has the highest expected value. At this point, the still uncertain decision maker can still perceive with a high degree of certainty what the ordinal relations are within each attribute. It seems perfectly plausible that a person might perform a noisy expected value calculation and still be able to determine with a high degree of certainty that one alternative ranks higher than another. To take an example from Chapter 3, a person might have a relatively noisy estimate of the area of two rectangles in millimetres, but have a very reliable perception of which of two rectangles is taller. Furthermore, research in vision has provided evidence that the
brain combines cues to depth even if they are correlated, e.g., linear perspective and texture gradients (Oruç, Maloney, & Landy, 2003). The fact that there is redundancy in this process likely reflects an attempt to make more robust predictions in the face of noise on both estimates.

The rational aspect of our approach is methodological and it is possible that other normative models may contribute separate insights in to the behaviour. Whilst we assume expected value maximisation in choices among prospects, Shenoy and Yu (2013) assume an inference of fair market value. Alternative approaches might yield other rational explanations. One alternative we previously considered was an information theoretic account of how having more samples in the target space (because the decoy is nearby) reduces the uncertainty in the estimate of the target’s true value, consequently making it a less risky alternative than the competitor. The aforementioned animal studies also show that irrational behaviour can still be interpreted as optimal (Fawcett et al., 2014; Trimmer, 2013). Another supportive stream of research concerns the Bayesian integration of priors reflecting experience of the statistical properties of the environment. Fennell and Baddeley (2012) for instance show that combining statements of probability with prior experience using Bayes can predict the deviations from Von Neumann and Morgenstern’s (1944) expected utility theory that are described in prospect theory (Kahneman & Tversky, 1979) as adaptive.

More recently there have been several critiques of Bayesian approaches to understanding cognition (Jones & Love, 2011; Bowers & Davis, 2012). Bowers and Davis (2012) argue that some Bayesian models of cognition are unfalsifiable due to flexibility in the priors they permit. However, as Hahn (2014) notes, priors (and other parameters) can simply be fit to environment or participant measurements leading to easily falsifiable predictions. Chapter 5 provides a case in point since the environmental priors were determined by analysing the statistics of the Wedell (1991) stimuli. The model could be even more rigorously tested by measuring a participant’s noise in expected value calculation which would leave no free parameters at all.

6.2.3 Objective 3: Derive Novel Predictions

The third question asked whether the model could make any novel predictions regarding the attraction effect. It is perhaps always possible to develop a rational account of behaviour by taking in to account enough bounds on cognition or assuming sufficiently ‘bizarre’ goals (Chater & Oaksford, 2000, p. 105). This does not necessarily mean that the inferences we make from such accounts are correct. An account may be taken more seriously if it can predict novel behaviours, it is to this end that we have sought to generate novel predictions from the model in Chapters 3 and 4.

The first of these predictions emerged entirely unexpectedly from our high level model
The model shows that the attraction effect configuration in the loss domain actually implies that the decoy should be perceived as the worst option of the three, whereas the target and competitor alternatives should be perceived to be unaffected by the presence of the decoy. This is the mirror image of the gain domain pattern, where the target is perceived to be better, and the competitor and decoy alternatives do not change.

The prediction is somewhat counter-intuitive. Many of the heuristic accounts described in Chapter 1 would suggest that the effect should be present in the loss domain. Take for instance Simonson (1989), according to this logic, participants choose the target in the gain domain because its dominance over the decoy provides them with a reason to justify their choice. In the loss domain there is no reason to assume that people would not also choose the target since it still strictly dominates the decoy. In a decision by sampling account of the attraction effect, the target counts the most ‘wins’ against the other alternatives and is therefore chosen more often. This explanation, should again be expected to hold in the loss domain.

The model not only predicts, but also provides an explanation for the effect - it is adaptive. Exhibiting a smaller attraction effect in the loss domain will result in smaller expected losses than would be the case if people exhibited the attraction effect to the same extent as the gain domain. If, as the model suggests, people are using the contextual information provided by the choice set to make better choices than they otherwise would. This means the attraction effect should be expected in the gain domain, but be significantly reduced in the loss domain. The evidence that we have gathered provides strong support for our claim that the attraction effect results from an adaptive strategy to make better decisions by taking the contextual information into account.

The identification of a causal rather than descriptive account of the attraction effect provides the benefit of being able to make predictions of novel phenomena (or their absence) in different environments. The mechanistic accounts of the attraction effect in the gain domain replicate the effect, but it is possible that they do not explain it. Our account suggests that far from being an inevitable side effect of the decision process resulting perhaps from lateral inhibition or range normalisation it is simply a behaviour that is beneficial in certain environments. When it is not beneficial we do not exhibit the effect, strongly suggesting that it does not occur inevitably as the result of mechanism but because it is adaptive in certain environments. This finding strikes at the heart of the distinction between the how and why explanations that are the product of developing descriptive mechanistic accounts and rational accounts respectively.

The results in Chapter 5 are subject to the same limitations as those of Chapter 3 given that the model is identical. There are however, some additional questions to consider with regard to the loss domain, not least do participants understand the task as it is intended? The data from Experiment 2 provide strong evidence that they did. It was possible to fit logistic regressions to participants’ data revealing that for a fixed level of risk (30%) peo-
ple increasingly chose the safer alternative as the value of the risky alternative decreased (larger losses). Likewise the data from Experiment 1 reveal that participants understood the ordinal relations consistently avoiding the decoy in both the gain and loss domains.

That decisions in the loss domain are different from those in the gain domain is well established in the decision making literature, not least in prospect theory (Kahneman & Tversky, 1979) which describes how people are more risk seeking in the loss domain. It is possible that in the loss domain people pay more attention given that losses loom larger than gains. If this is the case, perhaps people take more care in their evaluation of alternatives that involve losses. This would suggest less variance in their estimates of the expected value of each prospect and consequently a decreased likelihood of exhibiting the effect according to our model. Even if that were the process by which the attraction effect was reduced in the loss domain, the fact remains that our model suggests it is an advantageous process and therefore we have primarily provided an explanation of why people do not exhibit the attraction effect in the loss domain rather than a description of the decision process.

Perhaps more powerful evidence for the model would come from making entirely novel predictions (rather than reduction of an effect) and verifying these empirically. One potential avenue would be to consider the predicted decision effects that might result from other ordinal relations that could be constructed. The model could be applied to any set of ordinal relations among a varying numbers of alternatives. These might be analysed to predict new phenomena. Whilst this would be interesting, it suggests a return to contrasting behaviour with normative prescriptions in order for any predictions to be interesting. Whilst this was the starting point for this thesis, a more productive path from here on in might be to evaluate instead how, in general, the proposed model is an adaptive approach given all the possible ordinal combinations that can occur between sets of alternatives.

The modelling in this thesis has set out to demonstrate that the attraction effect is an adaptive behaviour. Consequently it was not designed to distinguish between, or compete with, existing models that describe processes or mechanisms that produce it. Nevertheless there is a striking empirical finding in the results of Chapter 5, and it is interesting to consider whether models such as MDF, LCA and AAM can simultaneously predict an effect in the gain domain and a much reduced effect in the loss domain. To the extent that these models focus on the differences between alternatives within attributes it seems likely that they would predict an attraction effect in the loss domain. This is conjecture though, and an obvious next step is to test which of these existing models predict that there should be an effect in the loss domain. This may be restricted in terms of which models can be applied in the loss domain, for instance initial attempts to model the context dependent advantage model (Tversky & Simonson, 1993) were difficult, since it is not clear how the authors would extend their model to the loss domain, nor how probability can be modelled.
as having negative values.

6.2.4 Summary

People generally adhere to the principle of value maximisation. Alternatives are assigned a value or worth and the highest is chosen. In certain restricted decision environments this process can leave a high degree of ambiguity as to which of the alternatives on offer is best (for instance when presented with gambles that have the same expected value). In these environments people make use of contextual information as provided by the other alternatives on offer to reduce their uncertainty about which alternative is best. Taking into account the environmental expected values implied by the ordinal relations in a choice set is one such way to do this. The result is that people make better decisions than they otherwise would.

The fundamental insight here is that the attraction effect can result in higher expected value decisions than a value maximising decision maker would make. This statement is only a contradiction if we do not take cognitive constraints into account. Once these have been factored into the decision problem IIA is no longer a condition of rational behaviour precisely because the decoy is no longer irrelevant, it provides information about the expected value of the target.

There are three levels at which the research reported here has contributed new insights. Firstly, the research has resulted in several novel empirical findings which can be considered separately from the extent that they support my interpretation of them. There is a novel finding that the attraction effect can be elicited in the rapid pointing task as shown in Chapter 2. In Chapter 3 the data show that the attraction effect is limited to choices between alternatives that are difficult to tell apart, and in Chapter 5 there is another novel finding that choices between prospects in the loss domain do not elicit the attraction effect.

A further contribution to knowledge are the analyses showing that the attraction effect may result from an attempt to achieve a more accurate estimate of the underlying expected value of an alternative. This analysis is, to my knowledge, the first to explain how an attempt to reduce uncertainty in the perception an alternative’s value can lead to contextual preference reversals.

From a more theoretical perspective, both the empirical data and the above analyses result from the novel application of the computational rationality approach to developing a causal explanation for challenging phenomena such as contextual preference reversals. This requires a rigorous understanding of what rationality means and how it can usefully be applied as a methodological tool in the cognitive sciences.
6.3 Conclusion

The attraction effect is an adaptive behaviour that results in better decisions than would be achieved were it absent. Understanding the attraction effect as beneficial resulted from using a computational rationality approach. This allowed us to ask what the best decision would be given a person’s cognitive capacities. Once these constraints have been factored into the problem it is possible to show why people exhibit the effect rather than be restricted to positing mechanisms that might reproduce it.

Explaining why, as well as how, a phenomenon like the attraction effect is exhibited is critical. It allows us to develop a deeper insight into the behaviour by establishing a causal link between cognitive bounds, environmental constraints and a person’s goal. The best evidence for this deeper level of insight, is that we were able to empirically verify a novel prediction that the attraction effect would be reduced in the loss domain. Finally, the approach has yielded exciting future research possibilities, not least that risk preferences might emerge from expected value maximisation with sensitivity to the statistical properties of the decision environment.
References


Herne, K. (1999, August). The effects of decoy gambles on individual choice. *Experi-


Experimental Psychology: General, 132, 3–22.


McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. Trends in cognitive sciences, 14, 348–356.


Noguchi, T., & Stewart, N. (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. Cognition, 132, 44–56.


Appendix A

Measuring the Attraction Effect

Each of the experimental chapters contains a detailed methods section. This appendix identifies and explains their common elements.

In all three empirical chapters our measure of the attraction effect follows a common method. Figure A.1 shows an attraction effect choice set containing three alternatives $A, B, D$. The literature on the effect tells us that adding a decoy option increases the likelihood that the resulting target option will be chosen. Suppose $A$ and $B$ are equally preferred such that each is chosen 50% of the time when only $A$ and $B$ are available. Adding the decoy to the choice set such that it is strictly dominated by $A$ will result in $A$ being chosen more often than $B$. In the configuration shown in the figure, $A$ might be chosen 60% of the time and $B$ 40% of the time. Note that $D$ is not chosen, but that it alters the ratio of choices $A$ to $B$. Now if we were to move the decoy option $D$ to the dashed area in the figure, $B$ might be chosen 60% of the time and $A$ 40% of the time.

The attraction effect is measured as the increase in choice probability that an alternative benefits from when a decoy is positioned to make it a target. In the above example, $A$ was chosen 50% of the time when there was no decoy, and 60% of the time when it was a target. Consequently the attraction effect would be reported as a preference reversal rate of 10%.

An alternative measure of preference reversal rate is not to use a base rate (when there is no decoy present), but to subtract the probability an option is chosen when it is the competitor from the probability the same option is chosen when it is the target. This simply doubles the size of the effect being measured. In the example above this would involve subtracting the rate $A$ is chosen when it is a competitor (40%) from the rate it is chosen when it is a target (60%), resulting in a preference reversal rate of 20%. This method was used by (Wedell, 1991) and is the measure we report in all of the empirical chapters.
Figure A.1: Measuring the attraction effect. Preference reversal rate was calculated as the rate A was chosen when it was a target minus the rate it was chosen when it was a competitor.

A.0.1 Between vs Within Measures

There are two ways preference reversal rate can be calculated. One is to give many participants one choice where an option is the target, and another where the same option is the competitor. A group level preference reversal rate can then be calculated. If 64% of participants chose option A when it was a target and 42% chose it when it was a competitor then the (group level) preference reversal rate would be reported as 22%. This procedure could be conducted as either a between or within subjects design.

The method we used in all of our experiments was to present each participant with the same choice set many times. For example, the configuration shown in Figure A.1 might be presented 10 times over the course of an experiment, allowing us to calculate a rate that option A is chosen from that choice set for each participant. In another 10 trials each participant would make choices from the same choice set but with B as the target (and by definition A as the competitor). We can therefore calculate the preference reversal rate for each participant as the rate A is chosen when it is the target minus the rate A is chosen when it is the competitor. For example, if Participant 41 chooses A 7 times out of 10 when A is a target and 5 times out of 10 when A is a competitor, his preference reversal rate is calculated as 20% (0.7 – 0.5). The preference reversal rates reported in the thesis for different conditions are aggregate data of preference reversal rates for each of the participants.
Appendix B

Expected Value

Expected value plays a large role in the thesis, both as an independent variable in our experiments and as a model of rational decision making. If a gamble is described as having \( X \) outcome values \( \{x_1, x_2, \ldots, x_n\} \) that occur with probabilities \( \{p_1, p_2, \ldots, p_n\} \) then the expected value of \( X \) can be calculated as

\[
EV(X) = \sum_{i=1}^{n} p_i x_i
\]

Expected value is the average return that would be expected from repeatedly playing a gamble. Thus a 20% probability of winning £50, otherwise £0, has an expected value of £10 since \( 0.2(50) + 0.8(0) = 10 \).

A rational, expected value maximising decision maker, when offered two prospects to choose from, would simply calculate the expected value of each and choose the prospect that had the larger expected value. Over the long run this strategy is guaranteed to achieve the highest return possible\(^1\).

It should be noted that the prospects we presented were always of the form \( p(x) \) or \( 1 - p(0) \) i.e., where a participant was offered ‘an 83% chance of $12’ it was clear that there was a complementary 17% probability of $0. This meant that the prospects could be easily presented as attraction effect stimuli such that one attribute was probability and the other was outcome value. The expected value of these prospects can be calculated as the product of the first two terms since the \( 1 - p \) outcome was always zero. This property had the advantage that the rectangle stimuli used in Chapter 3 could be described and evaluated in a mathematically identical way to the prospects. In both cases, the expected value of prospects and the area of rectangles could be calculated as the product of the attribute values. For instance, in Figure A.1 the areas described by the coordinates \( X_0, X_A, Y_0, Y_A \) and \( X_0, X_B, Y_0, Y_B \) could be used as target and competitor rectangle stimuli in the area judgement task.

\(^1\)Although for a bounded agent, see the model in Chapters 3, 4 and 5!