A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

http://wrap.warwick.ac.uk/144564

Copyright and reuse:

This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

Essays on context effects in behavioural economics

by

Anna Trendl

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Psychology

September 2019

THE UNIVERSITY OF
WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

Neil, Tim, I am immensely grateful for your excellent supervision, I have learnt so much from you. I would like to express my gratitude to Chris Starmer for making this studentship possible. I would also like to thank my family and my friends for their constant encouragement and the ESRC for their financial support. Finally, Tom, thanks for putting up with me.

Declarations

This thesis is submitted to the University of Warwick in support of an application for the degree of Doctor of Philosophy. It has been composed by myself and has not submitted in any previous application for any degree.

The studies presented here were carried out by the author in collaboration with Neil Stewart (Warwick Business School, University of Warwick) and Timothy Mullett (Warwick Business School, University of Warwick). All chapters were written by the author with feedback from the collaborators.

The concept of **Chapter 2**, and the design of the experiments were developed jointly. The experimental scripts were written by the author who also collected the data and fitted the models. The concept of the probability distribution approach was developed jointly, with Neil Stewart and Anna Trendl responsible for the technical implementation.

The concept of **Chapter 3**, and the design of the experiments were developed jointly. The algorithms to create the stimuli used in Experiments 1 and 2 were developed by the author, while the data collection was carried out by Neil Stewart. The analysis of the data was carried out by the author.

The concept of **Chapter 4**, and the design of the experiments were developed jointly. The algorithms to create the stimuli were developed by the author, who also collected and analysed the data.

The concept of **Chapter 5** was developed jointly. The data was provided by the West Midlands Police. The author designed and conducted the analyses.

Abstract

This thesis presents four research projects in the broader field of behavioural economics. In doing so, it aims to demonstrate the wide scope of behavioural economics as a scientific discipline, and illustrate its diverse methodological toolkit. Chapter 1 provides a broad overview of the birth of behavioural economics through the history of psychology and economics. Chapters 2–4 investigate context effects in decision making, while Chapter 5 describes an example of the profound effect visceral factors can exert on human behaviour.

More specifically, in Chapter 2, we investigate context-dependent choice behaviour within the framework of a popular cognitive model of decision making. Based on results from three tests using data from a value-based and perceptual choice experiment, we find that a mixture of absolute and relative valuation rules describes subjective valuation most accurately across two vastly different choice domains. In Chapter 3, we test whether the attraction effect, a well-known cognitive bias is present in choices involving naturalistic, complex stimuli. The results from two experiments suggest that the attraction effect does not extend to real-world choices. This finding serves as the basis for the research question we investigate in Chapter 4, where our aim is to explore the extent to which the strength of the attraction effect depends on the separability of the attribute dimensions. Potentially owing to the design of the experiment, the results of this investigation are mixed and inconclusive, which prevents us from providing a clear answer to our research question.

In contrast to the previous, laboratory-based experimental chapters, in Chapter 5, we use a large observational dataset to investigate the link between England's participation in national football tournaments and the number of reported domestic abuse cases recorded by the West Midlands Police. Using a regression approach, we find that alcohol-related domestic abuse increases by 61% following and England victory. As well as exploring the characteristics of this increase, we also re-analyse data from a previous study to reconcile our results with earlier findings from the literature, and demonstrate the robustness of the win effect. Our study is the first to highlight the instrumental role alcohol plays in the relationship between football and domestic abuse. Finally, Chapter 6 summarizes the findings, and discusses the limitations and possible directions for future research.

Chapter 1

A selective history of psychology and economics

1.1 Origins of modern psychology

Social and biological scientists have long been interested in exploring and understanding the factors shaping human behaviour. While the relative importance of nature versus nurture has been a subject of debate for centuries (Pinker, 2004), there is now a broad scientific consensus that human behaviour emerges as the result of the complex interactions between a range of external and internal factors, including genetic and environmental influences (Garcia Coll, Bearer, Lerner, Bearer, & Lerner, 2004). These factors have a profound impact on our personality, and exert a deep, often unconscious influence on our thoughts and actions. Modern psychology is primarily concerned with the scientific investigation of these factors, and aims to understand the pathways of their effect on human behaviour.

Psychology finds its roots in the philosophical traditions of ancient Western and Eastern civilizations and in the philosophy of the Enlightenment, but the birth of modern psychology as a scientific discipline is generally associated with the adoption of the experimental approach as a method of scientific inquiry in nineteenth-century Germany (R. Smith, 2013; Hergenhahn & Henley, 2013). Heavily influenced by the emerging scientific field of psychophysics, these early experimental investigations were primarily concerned with understanding perception and sensation (Mandler, 2007). In 1879, Wilhelm Wundt launched the first coherent experimental psychological research programme, which involved the measurement of choice reaction times and recording self-observations of sensations (Rieber & Robinson, 2001). However, with the subsequent rise of behaviourism in the twentieth century, the focus of experimental research in psychology has shifted from the mental experience to directly measurable behavioural responses to stimuli (R. Smith, 2013).

The cognitive revolution in the 1950s facilitated a dramatic paradigm shift in psychology, and ended the dominance of behaviourism. This transformative change was greatly enhanced by technological developments after the war, and the emerging fields of computer science and neuroscience equipped psychological scientists with new methodological tools to study the cognitive processes underlying attention, language, memory, perception, judgment and decision making (Miller, 2003). The nascent field of cognitive psychology followed the tradition of the early experimentalists from the 1800s in focusing on mental processes, but with strict scientific rigour (Leahey, 1979). In the 1970s, behavioural decision research, a new branch of cognitive psychology emerged, concerned with developing novel computational models of cognition to understand how cognitive limitations affect human decision making. Behavioural decision research specifically aimed to incorporate psychological insights within its theoretical models of decision making, in stark contrast with the dominant view on decision making in economics at the time (Angner & Loewenstein, 2012).

1.2 The separation of psychology and economics

The historical origins of modern economics and psychology are inextricably linked, and psychological states of the mind were central to the theories of early economic thinkers deciphering the forces governing human behaviour and driving economic activity. For example, in the eighteenth century, psychological insights and concepts were deeply embedded in the moral philosophy of Adam Smith (Ashraf, Camerer, & Loewenstein, 2005), and the psychological state of pleasure was the cornerstone of the utilitarianism of Jeremy Bentham and John Stuart Mill in the nineteenth century (Riley, 2008). Psychological hedonism was also a key concept in the utility theory underpinning the thinking of the members of the marginalist school of thought at the end of the nineteenth century, including William Stanley Jevons, Carl Menger and Léon Walras (Drakopoulos & Katselidis, 2017).

However, at the beginning of the twentieth century, mainstream economic theory went through a transformative change, called the "Paretian turn" (Bruni & Sugden, 2007). Vilfredo Pareto, who was heavily influenced by the positivist scientific thinking prevalent at the turn of the century, argued that economics should be a mathematical science, and that a theory of utility without any psychological concepts was possible and desirable. Psychological assumptions were deemed untestable and unscientific, and were feared to discredit economics as a science. As a result, mainstream economic thinking started to turn away from psychology, and neoclassical economists began to reformulate choice theory in order to free it from psychological ideas (Bruni & Guala, 2001).

These reformulated models, subsequently developed by John Hicks, Roy Allen and Paul Samuelson, postulated a rational, utility maximising decision maker with stable preferences (Drakopoulos & Katselidis, 2017). The view of the economic agent as a perfectly rational decision maker became the hallmark of mainstream economic theory, which underlies the most influential models in economic history developed at the time, such as the Arrow-Debreu general equilibrium model (Arrow & Debreu, 1954), and Expected Utility Theory (Neumann & Morgenstern, 1947).

However, the rationality assumption of mainstream economic theory faced growing criticism immediately from its birth in the first half of the twentieth century, and was widely deemed unrealistic by other social scientists (e.g., Bruni & Luigino, 2013; Sen, 1977). The most famous defender of the rationality assumption was economist Milton Friedman (Friedman, 1953), who took a truly instrumentalist position to shield the rationality assumption from criticism, and argued that the validity of a theory can only be tested through its predictive power, and not through the plausibility of its assumptions (Drakopoulos & Katselidis, 2017).

At the time of the cognitive revolution in psychology, a pioneering critic of the rationality assumption of economic theory was Herbert Simon, who used the term "bounded rationality" to describe a decision maker whose rationality was restricted due to both external (environmental) and internal (cognitive) limitations. His work is often referred to as the "old behavioural economics" (Sent, 2005). However, partly owing to the lack of a formalised approach to describe his insights, his theory failed to have a transformative impact on mainstream economics at the time, and had been subsequently misinterpreted in support of mainstream economic theory (Sent, 2005; Angner & Loewenstein, 2012).

1.3 Reunification: the birth of behavioural economics

The birth of behavioural economics as we know it today (the "new behavioural economics") dates back to the emergence of behavioural decision research in the 1970s, and was facilitated by the contributions of Paul Slovic, Baruch Fischhoff, Sarah Liechtenstein, Richard Thaler, and especially Amos Tversky and Daniel Kahneman (Sent, 2004). The development of Tversky and Kahneman's Prospect

Theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) marks the "return" of psychological insights into mainstream economic theory.

Prospect Theory is a descriptive, formalised theory of decision making under risk, which captures various, well-documented cognitive biases in choice. It includes deeply psychological concepts, in the form of reference-dependence, loss aversion, and non-linear probability weighting. The breakthrough of Prospect Theory within economics partly stemmed from the fact that it retained the highly formalised Expected Utility Theory framework to model violations of the rationality assumption, and thus did not represent a radical departure from the theoretical approach of mainstream economics (Angner & Loewenstein, 2012). The popularity of Prospect Theory has turned economists' attention to the abundant empirical evidence about "irrational" choice behaviour, and psychological concepts, such as regret, disappointment, self-control problems and social comparison processes were increasingly incorporated into economic models of decision making, whilst retaining the utility-based modelling approach of mainstream economics (e.g., Fehr & Schmidt, 1999; Thaler, 1981; Loomes & Sugden, 1982).

The popularity of Prospect Theory has established behavioural economics as a new discipline, and the 1980s saw the institutionalisation of behavioural economics as a separate academic discipline, with the foundation of the Society for the Advancement of Behavioral Economics in 1982, and the launch of two journals exclusively focusing on research in behavioural economics (Sent, 2004).

While the emergence of behavioural economics has highlighted the importance of employing an interdisciplinary approach to understand human judgment and decision making, other, significant findings from cognitive psychology only had a limited impact on the theoretical developments of mainstream economic theory. For example, certain concepts, such as the role of intuition in decision-making, and the wide range of heuristics and biases discovered by psychologists that often lead to optimal choices are difficult to adopt in a coherent utility-based model, which is still the standard approach to explain choice behaviour within economics (Kahneman, 2003; Gigerenzer, 2016).

1.4 The methodological toolkit of behavioural economics

Today, behavioural economics has an immense scope, and employs a vast range of methodological apparatus. The benefits of utilising an interdisciplinary approach to understand and predict human behaviour in various contexts is illustrated by the numerous proliferating subfields within behavioural economics, including neuroeconomics, behavioural finance, behavioural development economics, behavioural law and economics, and behavioural game theory (Angner & Loewenstein, 2012). Owing to its deeply interdisciplinary nature, behavioural economics has an eclectic methodological approach, and naturally borrows empirical methods traditionally used in closely-related fields, such as psychology, economics, and neuroscience (Angner & Loewenstein, 2012).

The popularity of the laboratory-based experimental method within behavioural economics is primarily inherited from psychology. It allows for investigations with high internal validity, while field experiments conducted in real-world settings derive insights about human behaviour in the "wild", with high external validity (Levitt & List, 2007). In identifying the external and internal factors influencing human behaviour in various real-world contexts, behavioural economics also heavily relies on the econometric analysis of observational real-world data, which often offers the opportunity to conduct a natural experiment (Angner & Loewenstein, 2012). Randomised controlled trials are widely used by developmental behavioural economists, and are indispensable methodological tools for evaluating the efficacy of various real-world interventions (Demeritt & Hoff, 2018). Insights delivered by these methodological approaches have informed public policy in many countries, and contributed to the development of various "nudges", non-regulatory interventions aiming to influence individual behaviour by changing the choice environment (A. Oliver, 2013).

The rapidly growing field of neuroeconomics focuses on the biological processes underlying decision making (Krajbich & Dean, 2015), and its methodological toolkit includes eye-tracking, functional magnetic resonance imaging (fMRI), and electroencephalography (EEG). To analyse the vast amounts of physiological data collected by these methods and accurately model the biological processes underlying this data, neuroeconomists heavily rely on computational process models of choice, an influential family of models originating from cognitive psychology (Fehr & Rangel, 2011; Krajbich, Oud, & Fehr, 2014).

A promising future direction for behavioural economics is the adoption of machine learning methods primarily used for the analysis of "Big Data". In the past two decades, the digital transformation has created ever-increasing sources of vast datasets generated by humans (e.g., through social media, search engines, financial transactions, location data generated by GPS), offering rich sources of information to study human behaviour (Blazquez & Domenech, 2018). While these datasets are typically too large to be analysed using traditional econometric methods, the rapidly expanding field of machine learning is focusing on the development of novel algorithms that can uncover generalisable patterns and structure in these vast datasets, and can be ultimately used for prediction tasks (Mullainathan & Spiess, 2017).

The research projects described in this thesis demonstrate the deeply interdisciplinary nature of behavioural economics, the broad variety of phenomena it studies, and the wide range of empirical approaches it employs. Across four research programmes, we study a variety of topics, including value sensitivity in perceptual and value-based choice, the boundary conditions for the attraction effect (an influential decision bias in behavioural economics), and the link between football and domestic abuse in the context of England. These research questions lie at the intersection of various disciplines, including cognitive psychology, neuroeconomics, and economics, and therefore serve as ideal research areas for behavioural economics.

In studying these research questions, we also demonstrate the wide variety of methodological approaches used in behavioural economics. In one chapter, we use eye-tracking data to fit a decision making model widely used in neuroeconomics. In another, we use a vast text corpus and a machine learning technique (natural language processing) to study the attraction effect. Finally, in contrast to these predominantly experimental investigations, in the last chapter we use a large-scale crime dataset and regression techniques to study the association between England football matches and the daily number of domestic abuse cases reported to the police.

1.5 Thesis overview

The following four chapters revolve around two main themes. Using experimental methods, Chapters 2–4 focus on violations of the rationality assumption in value-based choice, while Chapter 5 uses statistical analysis of a large observational dataset to explore an important example of the profound effects visceral factors can exert on human behaviour. Below is a more detailed description of each of the four chapters and the associated methodological approaches used.

Chapter 2 presents an extensive exploration of context-dependent choice behaviour, within the computational framework of a process model of decision making. Using data from two choice experiments with fundamentally different task domains (preferential choice task in Experiment 1, and perceptual choice task in Experiment 2), we assess the relative explanatory power of four distinct forms of context-dependent choice behaviour in three empirical tests. In the first test, we use eye-tracking data from Experiment 1 to implement a simulations-based model fitting approach, which, compared to previous investigations, represents a methodological improvement in several respects. In the second test, we describe an alternative, stochasticity-free method to derive choice probabilities.

The third, and final test is a qualitative comparison of the empirical performance of the four forms of context-dependency, using data from both experiments.

Chapter 3 also focuses on context-dependent choice behaviour, more specifically, it explores the boundary conditions of the most well-known and influential cognitive bias in decision-making research, the attraction effect. We test whether the attraction effect extends to choices with complex alternatives, as opposed to simplistic stimuli with numerical attributes so often used in experimental research. This research question was inspired by a previous debate in the literature, and drawing insights from this primarily methodological debate, we develop a methodology allowing us to conduct the first truly rigorous test of the attraction effect in choices with naturalistic stimuli, using data from two experiments. To create the stimuli, we use a machine learning technique, Latent Semantic Analysis (LSA), to establish the similarity of complex text objects.

Chapter 4 explores the research question arising from the results of Chapter 3. More specifically, our aim is to decipher the extent to which the attraction effect depends on the stimuli presentation. Drawing on earlier literature from the psychology of information processing, we design and conduct an experiment to test whether the strength of the attraction effect depends on the separability of the attribute dimensions.

While all chapters fall into the broad topic of understanding the factors shaping human behaviour, Chapter 5 departs from the laboratory-based experimental approach of the previous chapters. In this study, we use a large dataset covering eight years of crime data from the West Midlands Police to analyse the relationship between England's participation in national football tournaments (World Cup and European Championship) and the number of reported domestic abuse incidents. Given the random allocation of England match days, and the largely unpredictable outcome of each individual match, our crime dataset lends itself to a natural experiment allowing us to investigate the causal effect of the England national team's victory or loss on the number of reported abuse cases. Our study is the first to explore the role of alcohol in this relationship in the context of England. Using traditional econometric techniques, we present an extensive analysis of the various characteristics of the link between football and alcohol-related domestic abuse, and reconcile our results with seemingly contradictory results from earlier investigations.

Chapter 2

Investigating context-dependent preferences within the drift diffusion framework

2.1 Introduction

While there is ample evidence that our decisions are profoundly affected by the decision context, how exactly the perceived subjective value of an option changes as a function of the choice environment is still a matter of debate within the decision-making literature.

Over the past decades, several theoretical models of context-dependent valuation have been proposed. One of the earliest of these is Parducci's range-frequency theory (RFT; Parducci, 1963, 1965), a highly influential theory in psychophysics, offering an account of how objective stimuli get translated into subjective quantities. According to RFT, the perceived magnitude of the stimulus depends on two components, its range and rank value within the stimuli set. The range value of the stimulus captures its position with respect to the highest and lowest stimuli in the set, whereas the rank value reflects its position within the ordered set of stimuli. RFT has proven to be a powerful explanatory framework for context effects observed in a range of domains, such as strategic decision making (Vlaev & Chater, 2006), price judgments (Niedrich, Weathers, Hill, & Bell, 2009), and even pain perception (Watkinson, Wood, Lloyd, & Brown, 2013).

The range and rank principles have also been influential outside the RFT framework. For example, the effects of stimulus range on discrimination performance and magnitude judgments have long been the focus of perceptual psychology (e.g., Lockhead & Hinson, 1986). More recently, range effects have been increasingly incorporated into models emerging from the interdisciplinary field of decision neuroscience. These models mostly focus on explaining patterns of neural activity during decision making tasks in fMRI experiments. Given that neural firing rates are bounded from above due to biophysical constraints, it is a natural assumption that some sort of adaptive mechanism must take place to accommodate differences in the range of stimulus values that can be experienced in the real-world (e.g., Padoa-Schioppa, 2009; Soltani, de Martino, & Camerer, 2012; Rangel & Clithero, 2012; Louie, Khaw, & Glimcher, 2013).

The rank principle has also been successfully applied in several choice domains. For example, in an fMRI experiment where participants were shown pictures of monetary amounts they could win, Mullett and Tunney (2013) found that activity in certain brain regions reflected the current stimulus' rank position within the entire set of stimuli. In a socio-economic context, several studies have shown that the rank position of one's income within their respective social reference group (e.g., workplace, neighbourhood), but not the absolute level of income, is a significant predictor of a series of health outcomes, including mental health, self-reported happiness, and job satisfaction (e.g., Clark, Frijters, & Shields, 2008; Brown, Gardner, Oswald, & Qian, 2008; Boyce, Brown, & Moore, 2010; Daly, Boyce, & Wood, 2015).

The range and rank principles are two distinct approaches to value normalisation, where the transformed values are required to fall between 0 and 1. These can be contrasted with simpler forms of normalisation, where the values are divided by the maximum value in the set, which serves as a natural reference point for a subjective valuation scale. In the context of a choice experiment, this maximum can be represented by the highest value in the current choice trial (the local maximum), or, equally, it can be the highest value experienced during the entire experiment (the global maximum), which is equivalent to simply normalising all values by a constant.

In this research project, we were interested in comparing the relative explanatory power of these four different forms of subjective valuation (range, rank, local and global maximum normalisation). Our empirical strategy was to present participants with choice triplets in two experiments, in the form of complex preferential stimuli in Experiment 1, and perceptual stimuli in Experiment 2. The "objective value" of each of the three choice options in every trial could be quantified in both experiments, which allowed us to investigate how these four value transformation rules fare in predicting choice behaviour.

There exists a wide range of choice models in cognitive psychology, many of which could serve as an equally suitable theoretical framework to investigate context dependence in value-based choice. In this research project, we chose to employ the theoretical framework of a hugely influential cognitive model of choice, the drift diffusion model (DDM; Ratcliff, 1978) and one of its popular extensions, the attentional drift diffusion model (aDDM; Krajbich, Armel, & Rangel, 2010; Krajbich & Rangel, 2011), which incorporates the role of eye-movements within the standard DDM.

Considering a trinary choice context, our empirical strategy can be briefly described as follows. First, we identified four choice set manipulation rules (each representing a modification of the three items' values based on a given rule) for which these four value transformation rules (range, rank, local and global maximum normalisation) yielded different predictions regarding choice behaviour. All subsequent experimental stimuli in this research project were derived from these four choice set manipulation rules. Second, using data from a preferential choice experiment (Experiment 1), we obtained the best fitting parameter set for each participant and value transformation rule from fitting the aDDM to the choice and reaction time data. Using these best fitting parameter sets, we then compared the resulting log-likelihoods to identify the value transformation rule that best describes the choice behaviour of each participant. Finally, using data from the perceptual version of Experiment 1, in Experiment 2, we derived the best fitting parameters for each value transformation rule from fitting the DDM to group-level choice data, and conducted a qualitative comparison of the four value transformation rules' overall ability to predict changes in the choice proportions across the four choice set manipulation rules. The next section gives a brief overview of the DDM and its extension, the aDDM.

2.2 The drift diffusion model

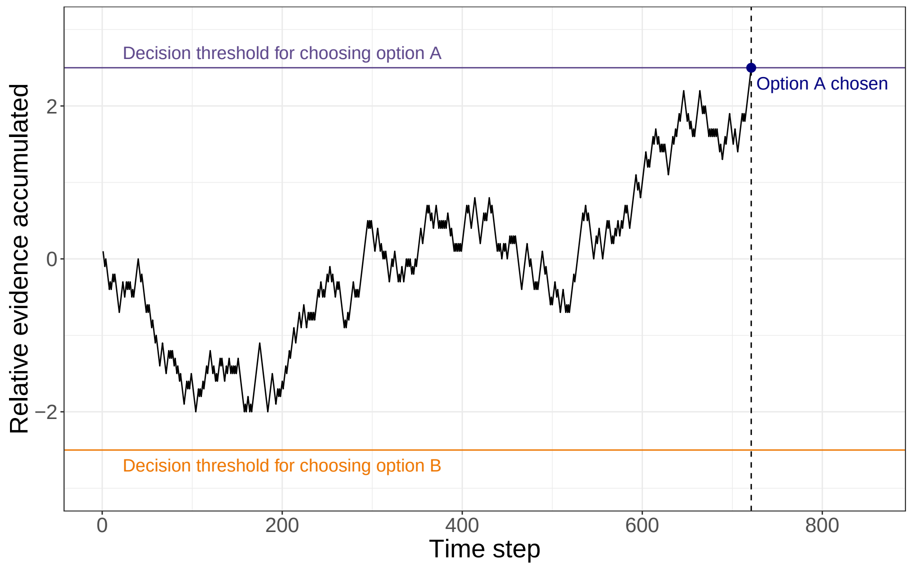
How do people choose when facing multiple options? What happens exactly during the choice process. What is the cognitive mechanism underlying the comparison of the alternatives? These are the questions process models of choice seek to answer. In cognitive psychology, one particularly influential type of process models is the family of sequential sampling models.

The overarching idea behind these models is that the evolution of preference in a given choice process is the result of noisy accumulation of evidence for each alternative throughout the decision process (hence the name "sequential sampling"), and a response is made when the accumulated evidence exceeds a certain threshold. There are many types of sequential sampling models, differing in whether the evidence is accumulated separately and independently for each option, and whether the threshold is defined to be absolute or relative (Teodorescu & Usher, 2013; Forstmann, Ratcliff, & Wagenmakers, 2016).

Of particular interest in psychology and cognitive neuroscience is a type of sequential sampling model that assumes a relative decision rule, and a separate and independent accumulation process for each choice option under consideration. In discrete time, this is modelled as a random walk process, whereas in continuous time it can be characterised by a diffusion process. The latter form is called the DDM (e.g., Ratcliff & Rouder, 1998; Ratcliff & McKoon, 2008), which is now the workhorse sequential sampling model in cognitive psychology.

In the drift diffusion framework, a key parameter of the model is the drift rate, which determines the average rate at which one of the thresholds is being approached during the choice process (Voss, Rothermund, & Voss, 2004). The drift rate also reflects the degree of similarity between the two options, and thus can be seen as a measure of task difficulty: when the options are very similar and the task is difficult, the process will have a low drift rate, and therefore it will take longer to reach one of the thresholds. Conversely, when the options are easily discriminable and the choice is “easy”, the drift rate will be high, and a threshold is reached faster. Figure 2.1 illustrates one such evidence accumulation process in the drift diffusion framework.

Figure 2.1: Illustration of the evidence accumulation process with options A and B.



While the DDM in psychology was first used as a model of memory retrieval (Ratcliff, 1978), it has since been applied to a wide range of choice task domains, including lexical decision making (Ratcliff, Gomez, & McKoon, 2004), numerosity discrimination (Leite & Ratcliff, 2011) and emotional processing (Mueller & Kuchinke, 2016). The popularity of the DDM in these various research areas stems from a number of factors.

First, and most importantly, the DDM both predicts reaction time distributions and choice probabilities remarkably well (Ratcliff, McKoon, & van Zandt, 1999; Forstmann et al., 2016). Second, it has been shown that manipulations of the decision task (such as changing task difficulty, accuracy motivation and reward structure of the task) correspond to expected changes in model parameters (drift rate, threshold distance and the starting point of the accumulation process respectively), indicating that the model is successful in capturing the cognitive mechanism underlying choice processes (Voss et al., 2004). Lastly, the model is inherently intuitive: it provides an elegant and plausible description of the evolution of preferences during the deliberation phase of decision making.

In addition to providing a psychologically plausible model of choice behaviour, the drift diffusion model has also been successfully applied in the field of cognitive neuroscience to explain both high- and low-level cognitive processes (Forstmann et al., 2016). Studies measuring decision-related neural activity in monkeys have found that over the course of a motion discrimination task, the firing rates of neurons in the lateral intraparietal cortex (LIP; an area in the brain responsible for attentional and decision-related processes, e.g. Shadlen & Newsome, 2001) exhibit a pattern closely resembling to the accumulation of noisy evidence (e.g. Churchland et al., 2011). In humans, studies using functional magnetic resonance imaging (fMRI) have identified distinct areas of the brain whose activation changes following value manipulations of model parameters (through changing the task design) in the DDM (for a review see Mulder, van Maanen, & Forstmann, 2014).

Owing to the popularity of the DDM, several extensions of the model have been proposed. One notable example is the incorporation of attentional processes in binary and trinary value-based decisions, known as the aDDM (Krajbich et al., 2010; Krajbich & Rangel, 2011). The aDDM modifies the standard DDM drift rate that governs the accumulation process based on the pattern of visual fixation during choice, resulting in a faster accumulation process for the option that is being fixated at a given time point. The mathematical formulation of the accumulation process in the trinary case with choice options left, centre and right, when option left is fixated is as follows:

$$E_t^{left} = E_{t-1}^{left} + d \cdot r^{left} + \varepsilon_t^{left} \tag{2.1}$$

$$E_t^{center} = E_{t-1}^{center} + \theta \cdot d \cdot r^{center} + \varepsilon_t^{center} \tag{2.2}$$

$$E_t^{right} = E_{t-1}^{right} + \theta \cdot d \cdot r^{right} + \varepsilon_t^{right}, \tag{2.3}$$

where E_t is the amount of evidence accumulated for a given option (the value of the accumulator corresponding to an alternative) up until time period t, θ is the penalty on the unattended items that can vary between 0 and 1 ($\theta = 1$ corresponding to the standard DDM model with no role of visual fixations, and $\theta = 0$ corresponding to the case of maximum penalty, such that the accumulation process is maximally dependent on the visual fixations), d is a constant that governs the rate of accumulation, r is the subjective value of the given option, and ε is the noise in the accumulation process. Therefore, for a given option, the total amount of accumulated evidence in each time period is given by the sum of the value of that accumulator in the previous time period, the value of the option discounted by the penalty on the unattended item (if it is not currently attended), and the noise component $\varepsilon \sim N(0, \sigma^2)$ (see Equations 2.1–2.3). The relative evidence position of each option at a given time period is then defined as follows:

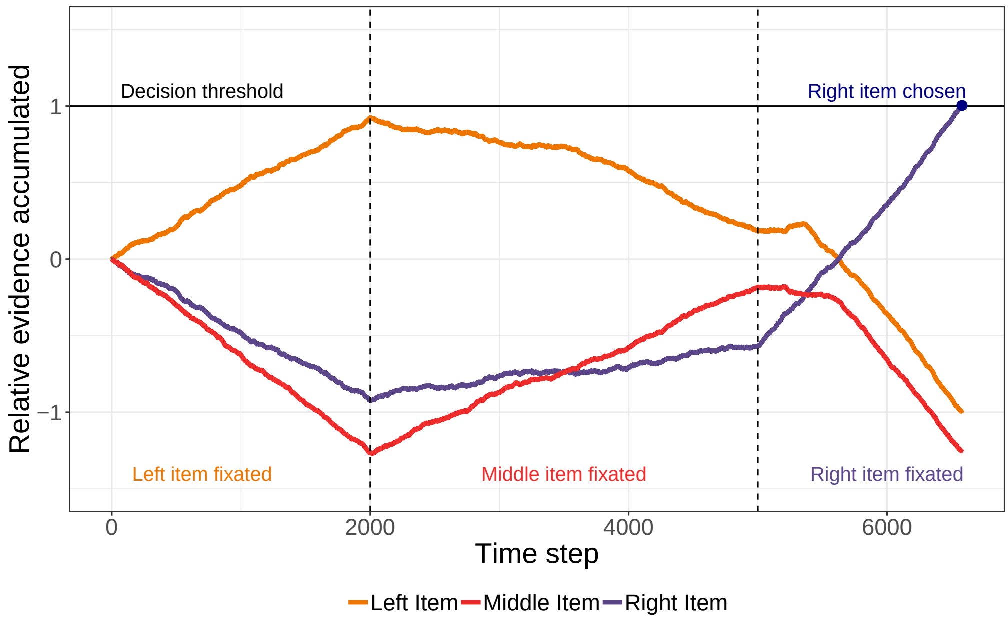
$$V_t^{left} = E_t^{left} - max(E_t^{center}, E_t^{right}) \tag{2.4}$$

$$V_t^{center} = E_t^{center} - max(E_t^{left}, E_t^{right}) \tag{2.5}$$

$$V_t^{right} = E_t^{right} - max(E_t^{left}, E_t^{center}) \tag{2.6}$$

As can be seen from Equations 2.4–2.6, the aDDM implements a next to best decision rule: at any given time step, each item's accumulator competes with the higher out of the other two. A choice is made when one of the relative evidence accumulators reaches a given threshold. The free parameters in the aDDM are θ, d and σ^2. Figure 2.2 illustrates how the accumulation process

Figure 2.2: Illustration of the evidence accumulation process in the aDDM framework ($r^{\text{left}} = 4$; $r^{\text{center}} = 3$; $r^{\text{right}} = 6$; $d = 0.0002$; $\theta = 0.3$; $\sigma^2 = 0.001$)



depends on the pattern of fixations in the aDDM framework. In general, through θ, the relative decision value for the fixated item increases, while it decreases for the unfixated items. All else being equal, varying the value of any of the other parameters also alters the accumulation process. For example, increasing d, the speed of integration results in a faster accumulation process for all options, reducing reaction times, while increasing σ^2 results in a noisier, less deterministic accumulation process. The speed of the accumulation also depends on the value of the options, r, such that the higher the option's value, the faster evidence accumulates for that option.

How does the aDDM improve the explanatory power of the DDM in understanding preferential decision making? Several studies have demonstrated the strong link between eye movements and subsequent choice behaviour. In a preferential choice experiment with pairs of faces used as stimuli, Shimojo, Simion, Shimojo, and Scheier (2003) have found that participants were more likely to fixate on the item they ended up choosing, a phenomenon called the gaze bias, which has also been observed in other eye tracking experiments involving preferential choice tasks (e.g. Armel, Beaumel, & Rangel, 2008; Bird, Lauwereyns, & Crawford, 2012). This effect tends to be much more pronounced towards the end of the decision: when retrospectively plotting how attention changes seconds before the choice was made, there is a much higher likelihood of the finally fixated item being the eventually chosen one too. This is called the gaze cascade effect (e.g. Shimojo et al., 2003; Glaholt & Reingold, 2009) or the late onset bias (Mullett & Stewart, 2016).

While these patterns provide strong evidence for the intimate link between attention and choice behaviour captured in the aDDM, the assumption that eye movements have a casual effect on preference formation has proven difficult to demonstrate. This difficulty stems from the fact that designing an eye-tracking experiment that mimics real-life choice scenarios, whilst allowing for exogenous manipulations of gaze in a non-invasive fashion is extremely challenging from a methodological point of view. Attempts to overcome this difficulty included explicitly controlling for exposure time by displaying only one item at a time (Armel et al., 2008), instructing participants to fixate on the item indicated with a specifically coloured frame in a binary choice task (Lim, O'Doherty, & Rangel, 2011), and prompting a decision once the target has been fixated for a pre-determined amount of time (Pärnamets et al., 2015). While these studies have demonstrated that items that were fixated longer were also more likely to be chosen, the artificial nature of these choice scenarios mean that they do not provide unequivocal evidence that eye movements have a causal effect on preferences in real-life choice scenarios.

However, in a recent study, Gwinn, Leber, and Krajbich (2019) directly tested the causal effect of eye movements on preferences. They used probability cueing, an elegant, non-invasive attentional learning technique, which induces attentional biases that spill over to a subsequent choice task. Their results from four studies lent support to the idea that attention has a causal effect on choice behaviour, providing the strongest evidence for this relationship yet.

While the causal role of fixations in choice behaviour is not a critical assumption in the aDDM framework (Krajbich, 2019), the model can accommodate both the gaze bias and gaze cascade effect. First, fixating on one item for longer means that the amount of relative evidence accumulated in favour of this item will be higher, therefore it is more likely to be finally chosen (gaze bias effect). Second, as evidence accumulation is faster when an item is being fixated, it is more likely that the item's relative evidence accumulator reaches the threshold during a fixation on that item (gaze cascade effect).

In addition, a comprehensive investigation of six eye-tracking studies found that attention amplifies the underlying value of the option under consideration, as opposed to providing a value-independent boost in the form of a constant added to the accumulation equation, supporting the multiplicative formulation of the aDDM (S. M. Smith & Krajbich, 2019). Furthermore, comparisons of aDDM predictions and actual eye-tracking data have shown that the aDDM provides a remarkably good fit to fixation, reaction time and choice data (Krajbich et al., 2010; Krajbich & Rangel, 2011).

Owing to its high explanatory power, we decided to investigate context-dependent choice behaviour within the aDDM framework. When fitting the aDDM to actual choice data, the value of the options (and consequently the drift rate) is usually defined as the item ratings given by participants in a rating task that precedes the choice task (r^{left}, r^{middle}, and r^{right} in Equations 2.1–2.3). These ratings are assumed to reflect the "objective value" of an option, evaluated independently from the other options available. In this research project, our aim is to incorporate the idea of context-dependent valuation (in the form of the range, rank, local, and global maximum value transformation rules) within the aDDM framework through these item ratings.

In other words, we compare four competing accounts of how the collected evidence samples are defined within the model framework. In the original model, it is the absolute values of the options (preference ratings on a scale from -10 to 10; e.g., Krajbich & Rangel, 2011) that govern the drift rates, whereas in our formulation, it is the normalised, subjective values of the options that feed into the accumulation equation at each time step. Importantly, we did not explicitly modify how the options are contrasted and integrated within the model (described in Equations 2.4–2.6).

2.3 Overview of Experiment 1 and 2

To test the relative explanatory power of various forms of context-dependent valuation within the aDDM framework, we conducted two experiments. Experiment 1 was a preferential choice experiment, where participants were presented with choice triplets created from 100 movie posters, while their eye movements were recorded. In a separate rating stage that preceded the choice task, we obtained independent preference ratings from each participant on the set of 100 movies. We investigated how preferences change as a function of the choice context by using these preference ratings as inputs to the four value transformation rules. In Experiment 2, we were interested in the same question, but we used perceptual stimuli, in the form of rapidly updating sequences of numbers drawn from a normal distribution with a fixed mean (taken from Tsetsos, Chater, & Usher, 2012). Analogously, we used these fixed means as inputs to the four value transformation rules.

To compare the explanatory power of the four value transformation rules, we used two fundamentally different methodological approaches. First, using data on eye movements during choice from Experiment 1, we obtained the best fitting set of aDDM model parameters for each valuation rule and participant, and based on the resulting log-likelihoods, we determined which valuation rule describes best each participant's choice behaviour. In order to obtain these best fitting parameters, we used two alternative methods to estimate the choice probabilities: a traditional simulations-based method, and a novel probability distribution method, which circumvents a methodological problem arising from simulating out a stochastic model. Second, using a DDM simulations approach and choice data from both experiments, we compared the choice proportion predictions for each value transformation rule, which served as a further test of the explanatory performance of each rule.

Our contribution is twofold. First, we conduct a rigorous comparison of the performance of the four subjective transformation rules, delivering insights about context-dependent choice behaviour across multiple stimuli domains. Second, we develop a novel probability density evolution approach for fitting the aDDM to eye-tracking data.

2.4 Subjective Transformation Rules

We considered the four value transformation rules detailed in Section 2.1: the range, rank, local, and global maximum rule. In the context of a trinary choice experiment, each rule takes in three "objective values" x_1, x_2, x_3 (these are the "raw" independent preference ratings for each movie in Experiment 1, and the mean of each distribution from which the rapidly updating sequence of numbers are drawn in Experiment 2), and transforms these into normalised "subjective values" (required to fall between 0 and 1).

Range Rule. The range value follows from the RFT and reflects the perceived distance of the stimulus from the stimulus with the lowest value, as a proportion of the overall distance between the highest and lowest valued stimulus in the stimuli set. It captures the idea that valuation is sensitive to the highest and lowest value encountered (or, equivalently, to the overall range of values) in a given choice set. Formally, this can be described by the following equation:

$$Range_i = \frac{x_i - min(x_1, x_2, x_3)}{max(x_1, x_2, x_3) - min(x_1, x_2, x_3)} \tag{2.7}$$

Note that the items with the lowest and highest original values are always assigned 0 and 1 respectively, while the middle item's value can vary between 0 and 1.

Rank Rule. The rank rule also follows from RFT, and reflects the "frequency" of the stimulus, which can be translated as its ranked ordinal position in the choice set. The rank rule reflects a simple mechanism of valuation, whereby the subjective value of an item is solely determined by the number of items with lower or higher values in the choice set, and the extent of the difference between item values does not matter. Formally, given an ordered set of three stimuli x_1, x_2, x_3, the rank values of the stimuli are given by

$$Rank_i = \frac{i-1}{n-1} \tag{2.8}$$

Note that according to the rank rule, the three subjective values assigned to the items with the lowest, middle and highest value are always 0, 0.5 and 1, respectively.

Local Maximum Rule. The local maximum rule generates subjective value representations by dividing the value of each stimulus by the highest value encountered on that trial, and in this way it corresponds to the normalised version of the raw ratings used in Krajbich and Rangel (2011). This reflects a bias towards the most rewarding, and hence potentially most salient option in the current trial. Formally, the subjective values of a set of stimuli according to the local maximum rule is given by

$$Localmax_i = \frac{x_i}{max(x_1, x_2, x_3)} \tag{2.9}$$

Note that the stimuli with the highest value is always assigned a subjective value of 1, while the other two values can vary between 0 and 1 (exclusive).

Global Maximum Rule. The global maximum rule is identical to the Local Maximum Rule, except that the objective values are normalised using the highest value encountered in the whole of the experiment (and not just on that trial). This means that all values are normalised by the same constant, therefore, the global maximum rule is equivalent to the original version of the aDDM (e.g., Krajbich & Rangel, 2011), and can serve as a natural reference point to test whether the explanatory power of the model can be increased with a different value transformation rule. Formally, in an experiment where the items that can be encountered are $x_1, x_2, ..., x_n$, the subjective value of a stimuli set according to the global maximum rule is given by

$$Globalmax_i = \frac{x_i}{max(x_1, x_2, ..., x_n)} \tag{2.10}$$

Note that when an item in the current trial has the highest value in the whole of the experiment, this rule is equivalent to the local maximum rule, otherwise the values can vary between 0 and 1 (exclusive).

2.5 Choice Set Manipulations

After identifying the four value transformation rules detailed above, the next step was to find a set of choice set manipulations for which these transformation rules (accommodated within an aDDM or DDM framework) produce different predictions about choice behaviour. A choice set manipulation represents some modification of the three options' objective values based on a certain rule (e.g. doubling each value, or adding a constant to each of them). The key idea is that the four subjective value transformation rules (range, rank, local and global maximum) differ in their predictions about how these choice set manipulations affect choice proportions, which allowed us to assess their relative empirical performance.

We chose to investigate the effect of the following four distinct choice set manipulations: adding the same constant to each of the three values, multiplying each value by a constant, and having a distant or a close second, and third value.

Figure 2.3: Subjective values by choice set manipulation and value transformation rule. The four columns show each choice set manipulation, while the rows correspond to the four value transformation rules.

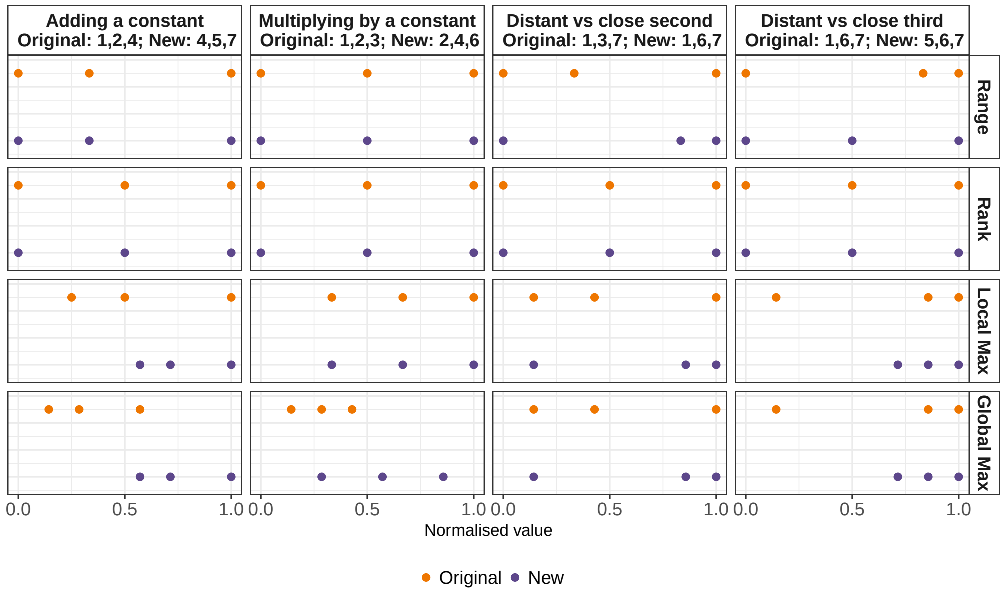


Figure 2.3 shows the subjective values (these are required to fall between 0 and 1) derived from each transformation, assuming a raw value scale from 1 to 7. The orange dots are the original values, while the purple dots show the new, transformed values. For example, in the case of adding a constant (first column of Figure 2.3), we added 3 to the original value set 1,2,4, which resulted in a new value set 4,5,7, whereas multiplying 1,2,3 by 2 resulted in the new value set 2,4,6 (second column). As illustrated by the four rows of Figure 2.3, the subjective values of the original and new value sets often differ across the four value transformation rules, resulting in different predictions about how these choice set manipulations affect choice behaviour.

For example, when adding a constant to a set of values, the range and rank transformed values are unaffected, while the global maximum transformed values only shift by a constant, therefore these three rules predict no change in choice behaviour. However, under the local maximum rule, the distance between the best and the other two values narrows, which is predicted to increase the relative choice share of the middle and worst options at the expense of the best option. When values are multiplied by a constant, only the global maximum rule predicts any change, namely, that as the distance between the three options increases, the best option becomes relatively more attractive compared to the other two. When increasing the value of the middle option while holding everything else constant (distant vs close second), the range and both maximum rules predict that the relative share of the middle option increases at the expense of the best option, while the rank rule predicts no change. Finally, the two maximum rules predict that increasing the value of the worst option while holding everything else constant (distant vs close third) will increase the share of the worst option at the expense of the middle option, the range rule predicts that it will increase the share of the best option and decrease the share of the middle option, while the rank rule predicts no change. Table 2.1 summarizes these predictions.

Table 2.1: Summary of the predicted effect of each choice set manipulation, by value transformation rule

		Adding a constant	**Multiplying by a constant**	**Distant vs close second**	**Distant vs close third**
Range	Best	NC	NC	Decrease	Increase
	Middle	NC	NC	Increase	Decrease
	Worst	NC	NC	NC	NC
Rank	Best	NC	NC	NC	NC
	Middle	NC	NC	NC	NC
	Worst	NC	NC	NC	NC
Local Max	Best	Decrease	NC	Decrease	NC
	Middle	Increase	NC	Increase	Decrease
	Worst	Increase	NC	NC	Increase
Global Max	Best	NC	Increase	Decrease	NC
	Middle	NC	Decrease	Increase	Decrease
	Worst	NC	Decrease	NC	Increase

*NC = No Change

Our original plan was to present participants with these seven distinct trinary choice sets (124, 457, 123, 246, 137, 167, 567) to maximise the discriminatory power of our experiments. However, due to an experimenter error, in

Experiment 1, the displayed choice options were quasi-random, resulting in 84 unique choice sets instead of the originally planned seven. We nevertheless proceeded with fitting the aDDM to the data from Experiment 1, and had enough data for the relevant seven choice sets to conduct a further comparison of choice behaviour in Experiment 1 and 2.

2.6 Experiment 1

2.6.1 Method

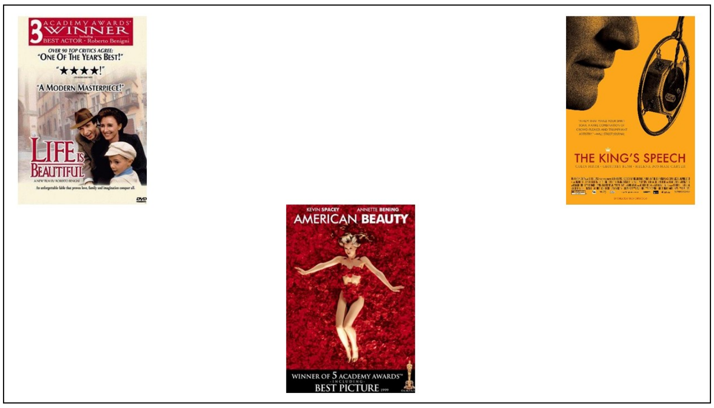
Experiment 1 was a preferential choice experiment, where participants' eye movements were recorded during choice. Based on previous eye-tracking studies (e.g., Krajbich et al., 2010; Krajbich & Rangel, 2011), we decided in advance that a sample of 50 participants should provide enough statistical power for our purposes. To reach this sample size, we recruited sixty participants overall, out of which ten participants could not be eye-tracked. Participants were recruited through the University of Warwick's Research Participation System, and were paid £7.00 for taking part in the experiment. We did not record data on participants' gender, as we did not expect it to affect the results. Ethical approval was obtained from the Department of Psychology, University of Warwick.

The experiment consisted of two parts. First, participants who signed up for the study were required to complete an online questionnaire, where they had to rate 110 movies on a scale from 1 to 7, based on how much they would like to see the given movie (1 being not at all, and 7 being very much so). We also asked them whether they have seen the movies before. The rating task began with 10 test trials (these were the same movies for all participants), so that participants could familiarise themselves with the task, and the type of movies they will be required to rate. Then, they were informed that the actual rating task is about to start. During this rating task, participants were presented with the 100 movies in a randomised order, and we instructed them to try to use the whole range of the scale as much as possible. This was important, because the second part of the experiment relied on these ratings, and could not be run without at least one movie for each rating. Some participants failed to do this, and so they were asked again to complete this task before coming to the lab for the second part of the experiment. Figure 2.4 shows an example trial from this task, which took about 15 minutes on average.

Figure 2.4: Example rating task from the first part of Experiment 1.

Those who successfully completed the first part were invited to the lab for the second part of the study. In the second part, participants were presented with 100 trinary sets of the pre-rated movies, and were asked to pick the one they would like to see the most on each trial by pressing the corresponding keyboard key (left, down, right arrow), while their eye movements were recorded at 500 Hz using an EyeLink 1000 (SR Research). The order of the choice triplets and the display order of the three movies in a given trial were both randomised. The eye tracker was calibrated after every 25 choices (four times overall over the course of the experiment), and each trial began only after participants fixated on a centrally presented fixation cross for 500 ms. We did not add a text prompt to the screen to avoid additional visual distractions. Figure 2.5 shows an example trial from the second part of this experiment.

Figure 2.5: Example choice task from the second part of Experiment 1.



2.6.2 Stimuli

We chose 110 movies that received the most votes in 10 distinct genres (adventure, comedy, drama, family, fantasy, history, mystery, romance, sci-fi, thriller) on IMDb (the most extensive movie information database) as of March, 2016. In the first part of the experiment, the first ten trials were practice trials (one movie from each genre), which aimed to give participants an idea about the range and type of movies they will be asked to rate.

In the second part of the study, the displayed trinary sets were originally generated based on the participant's ratings to reflect the choice set manipulations. However, as mentioned above, the choice sets displayed during the experiment were unintentionally quasi-random, due to the incorrect numbering of the 110 stimuli pictures. Each number still corresponded to a unique movie (albeit this assignment was random), but the choice sets, which were based on the movie numbers, were displayed as planned. This resulted in 84 unique trinary choice sets, defined by the raw rating of each of the three options, as opposed to the originally planned seven described in section 2.5.

In order to avoid the same set of movies appearing again, the script that generated the choice triplets aimed to ensure that all the stimuli with a given rating got used before any repetition of stimuli occurs. While we tried to vary the stimuli appearing as much as we could, repetition was inevitable in some cases, since a few participants gave very uneven ratings (e.g., gave a rating of 1 to only a few movies).

Given that our main aim was to fit the aDDM to the data to compare each value transformation variant, we decided it was still worth proceeding with the analysis with the current dataset (even though the data now contained considerably fewer trials that could help us delineate the differences in the explanatory power of the four value transformation rules).

2.6.3 Exclusion criteria

As mentioned in section 2.6.1, we had usable eye-tracking data from 50 participants. For a few participants, the calibration quality was not satisfactory (defined as "poor" by the EyeLink system) in some parts of the experiment, and therefore we excluded these trials, which accounted for about 2.5% of all fixations. We also excluded the first three trials of each participant, and each trial right after a calibration, to make sure we only included fixations that were part of the choice process. We then excluded fixations that fell outside the three areas of interest, and trials in the upper 1% of the reaction time distribution. Finally, we only kept trials where all three of the choice items were fixated at least once. After applying these exclusion criteria, the average number of trials per participant was 80.

2.7 Fitting the aDDM

Using data from Experiment 1, we wished to compare the four value transformation rules' explanatory power within the aDDM framework, on the basis of their respective log-likelihoods of producing each participant's data. To do this, we needed to establish a method for deriving choice probabilities, given a specific set of parameters. We used two different methodological approaches to derive these probabilities, a simulation-based approach, and a probability distribution approach. Below we first describe the standard model fitting approach used in the aDDM literature, and then we highlight how our simulation-based empirical strategy builds on, and improves this standard approach.

When fitting the aDDM, the common strategy is to fit the model on the group-level, and simulate out the choice process for all parameter value set (θ, the penalty on the unattended item, σ, the noise parameter, and d, the speed of integration) and preference rating combinations, with fixations sampled from the empirical distribution of fixations, conditional on the preference rating difference (e.g., Krajbich et al., 2010; Krajbich & Rangel, 2011; Fisher, 2017).

Then, given a discrete number of reaction time bins, the number of simulations falling into each choice and reaction time bin can be calculated for all parameter and rating difference combinations, and the probability that a simulation trial finishes in a given choice and reaction time bin can be derived. To obtain the log-likelihood of the data for each parameter combination, these probabilities are multiplied by the number of empirical data points falling into each choice and reaction time bin, and their logarithm is taken and summed up. One or two iterative grid searches can be carried out using this estimation method, resulting in the final estimate of the best fitting parameter values, and the corresponding log-likelihood.

Our empirical approach differs from this method in several respects. First, as opposed to a group-level approach, we derived the best fitting parameters for each subject to account for individual heterogeneity. Therefore, our primary aim was to find the best fitting parameter set and the corresponding log-likelihood value for each participant and transformation rule (of which there were 200, given 50 participants and four subjective value transformation rules) to determine which transformation rule yields the best fit for each participant. Second, to derive the most accurate estimate of the predicted probability of each trial falling into a given choice and reaction time bin, we used the eye movement pattern on that very trial (as opposed to probabilistically sampling it from an empirical distribution, conditional on the rating difference). Third, in addition to the grid search, we also used a Nelder-Mead optimization algorithm to find the best fitting parameter values for each participant and value transformation rule. Finally, we used 100,000 simulations to lessen the effect of noise (an inherent feature of the aDDM) and obtain more reliable choice probability estimates (previous studies used between 1000 and 4000 simulations, e.g., Krajbich & Rangel, 2011; Fisher, 2017).

In addition to these improvements, to evaluate the robustness of our estimates as well as to circumvent a methodological issue stemming from applying a traditional simulations-based approach to a stochastic model, we also used a novel probability distribution-based approach to derive the best fitting parameter combinations and corresponding log-likelihoods for each participant and value transformation rule.

Before the parameter estimation, we discretized the eye movement data. To do so, we divided each trial into 200 ms bins, where each bin was one step in the evidence accumulation process assumed by aDDM. The fixated item in a bin was sampled probabilistically, based on the proportion of time spent on fixating on a given item in each bin. For example, if in one bin, 50 ms was spent on fixating on item 1, and 150 ms was spent on fixating on item 2, then in the simulations item 1 had a 25%, while item 2 had a 75% probability of being chosen as the fixated item in that bin.

When fitting the aDDM for each value transformation rule, we used the transformed ratings in place of r described in Equations 2.1–2.3 in section 2.2. However, the unintentionally wide variety of value triplets displayed in Experiment 1 posed a question about how to derive the range transformed values in case there were two or three identical values in the choice set (it is straightforward in the case of the two maximum rules and the rank rule). We decided to code the transformed range values as 1 if all values were identical. If there were two equal values in a triplet, they were either coded as 1 or 0 (depending on whether the third value was lower or higher than the two identical ones, respectively).

2.7.1 Simulation-based method

To summarise the main novel aspects of our simulation approach, we model individual trials from individual participants (as opposed to deriving group-level parameter estimates), and use the eye movements specific to the very trial we are modelling (as opposed to randomly drawing it from the distribution of fixations from all trials). In addition, we find the best fitting parameter estimates through an optimization process, where the probability estimates are based on 100,000 simulations of each trial.

To obtain the relevant parameter sets for each participant and value transformation rule, we first simulated out each trial, and calculated the probability that the eventually chosen item was selected in the correct time bin. Note that there are two sources of variation in these simulations: the inherent noise in the accumulation process (captured by σ), and the non-deterministic fixation pattern arising from the transformation of the fixation data into time bins.

To find the best fitting parameter sets, we started off with a grid search with the following values: $\theta = \{0.2, 0.45, 0.6, 0.85, 1\}$, $\sigma = \{0.1, 0.14, 0.2, 0.28, 0.4\}$, $d = \{0.1, 0.17, 0.31, 0.56, 1\}$, 125 parameter sets overall, for each participant and value transformation rule. These parameter values were chosen after an iterative process, ensuring that most participants' best fitting values fell in the middle of the range of the possible grid values. This was done to ensure that the algorithm does not have to move considerable lengths in the parameter space to find the best fitting parameter set in the subsequent optimization process.

Upon obtaining the best fitting parameter set from the grid, we started a Nelder-Mead optimization algorithm with a maximum iteration number of 200, using the resulting parameter set from the grid search as starting points for each participant and value transformation rule. The same process was repeated once more using the results from the first Nelder-Mead optimization process as a starting point, to increase the probability that we find the parameter set that results in the highest log-likelihood of producing the data.

However, there is a methodological difficulty arising from the stochastic nature of the aDDM. Due to the inherent noise in the process, two runs of simulations of the same trial with the same parameter values naturally gives slightly different results. This is problematic if the estimated probability of the trial is generally low (near zero), as it can happen that in one simulation run, the estimated probability is a positive number, while in the next one it is zero. When taking the logarithm of all trial probabilities and summing them up, having just one trial with a zero probability results in a log-likelihood value of minus infinity. At the next simulation attempt, it is possible that the estimated probability for the same trial will be positive this time, and depending on the estimated probability of the rest of the trials, this can in turn result in a relatively high overall log-likelihood value. Such dramatic changes in the log-likelihood values can easily lead to the Nelder-Mead process getting stuck at a given point in the parameter space.

This is because the Nelder-Mead optimization method was initially developed for deterministic models, and it has been demonstrated that substantial noise in the underlying model can lead to false convergence (e.g., Chang, 2012; Barton & Ivey, 1996). While running the second Nelder-Mead did not lead to substantially different parameter estimates for most participants and value transformation rules, we still wanted to alleviate concerns about the validity of the simulations method, and decided to compare the results from the simulations approach with those obtained from an alternative, deterministic implementation of the aDDM.

2.7.2 Probability distribution method

To circumvent the problem arising from the stochastic nature of the aDDM, we utilised an alternative estimation method that is free of stochasticity. Rather than deriving the probability of the data through noisy simulation, we calculate it directly, by using a numerical evidence integration technique. This approach removes the problem of variable log-likelihood values, thus making the parameter search much easier.

In short, our approach was to derive the probability distribution of the accumulation process over the relative evidence states ($E_1 - E_2$ and $E_1 - E_3$ from Equations 2.1–2.3 in section 2.2, with $left, centre, right$ changed to 1, 2, 3, respectively), for each time bin and parameter combination. This method allowed us to directly calculate the exact choice probability for each trial in our data. Our approach is conceptually equivalent to that used by Diederich and Busemeyer (2003), who approximated the diffusion process using Markov Chain Theory, and derived the transition probabilities between different evidence states. However, we apply this method to a more complex setting than described in Diederich and Busemeyer (2003), in that implementing the aDDM entails a trinary choice case, with variable drift rates within a trial (stemming from changes in attention).

We chose to base our calculations on a "best versus average of other two" termination rule:

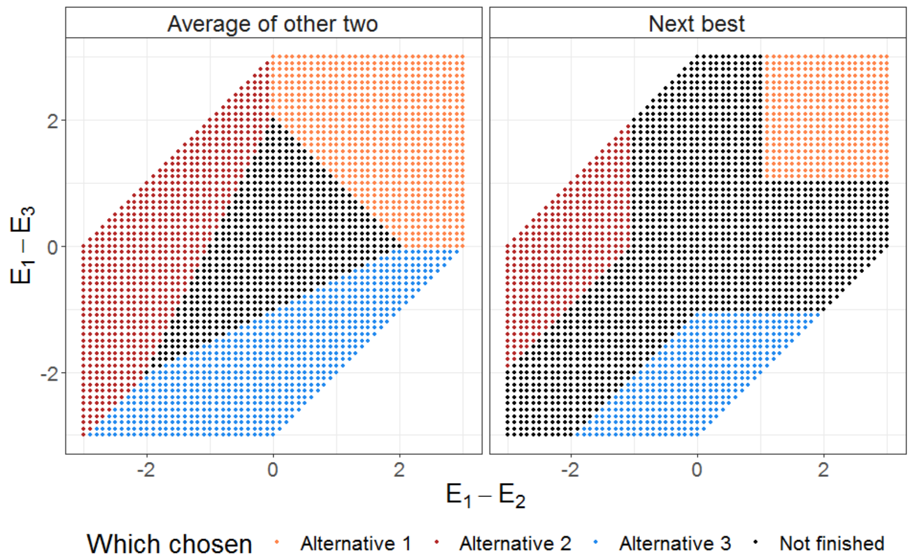
$$V_t^1 = E_t^1 - \frac{E_t^2 + E_t^3}{2} \tag{2.11}$$

$$V_t^2 = E_t^2 - \frac{E_t^1 + E_t^3}{2} \tag{2.12}$$

$$V_t^3 = E_t^3 - \frac{E_t^1 + E_t^2}{2} \tag{2.13}$$

as opposed to the original "best versus next best" formulation seen in Equations 2.4–2.6 in section 2.2. However, we did not expect this modification to have a dramatic effect on our results, since previous research had shown that the two rules yield very similar predictions (Krajbich & Rangel, 2011). The reason why we preferred the "best versus average of other two" rule over the "best versus next best" rule is because it results in a bounded area of the finished evidence states (corresponding to combinations of E_1, E_2 and E_3, where the threshold had been reached, and one of the options had been chosen), defined in terms of the two relative evidence states ($E_1 - E_2$ and $E_1 - E_3$). This is illustrated by Figure 2.6.

Figure 2.6: Finished evidence states with Average of other two and Next best termination rules.



As mentioned, we wished to calculate the probability distribution of the accumulation process over the relative evidence states $E_1 - E_2$ and $E_1 - E_3$ (each of which is represented by a dot on Figure 2.6), for a given parameter set and time step. Assuming that each step in the process can be described by a two-dimensional vector with components

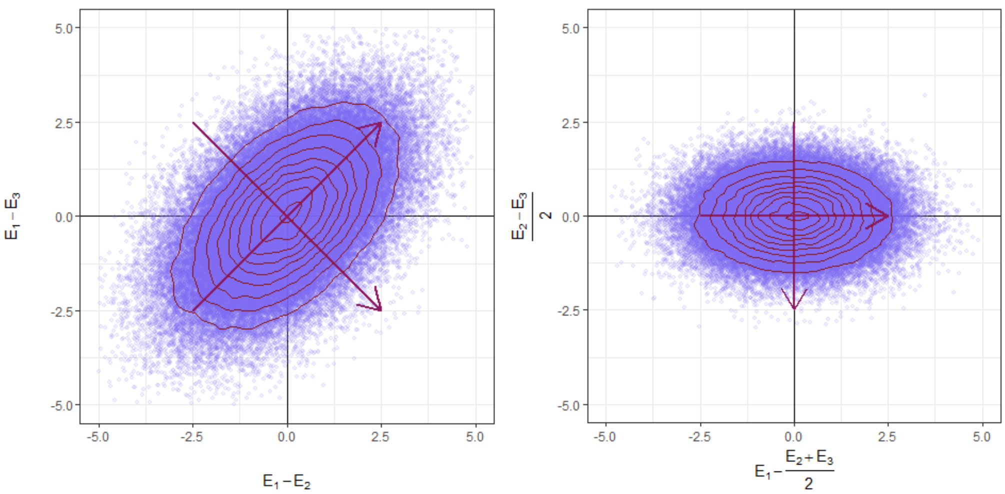
$$
\begin{aligned}
x &= E_1 - E_2 = d(\theta_t r^1 - \theta_t r^2) + (\varepsilon_t^1 - \varepsilon_t^2) \\
y &= E_1 - E_3 = d(\theta_t r^1 - \theta_t r^3) + (\varepsilon_t^1 - \varepsilon_t^3),
\end{aligned} \tag{2.14}
$$

the covariance matrix between the two axes is given by

$$
COV = \begin{bmatrix} 2\sigma^2 & \sigma^2 \\ \sigma^2 & 2\sigma^2 \end{bmatrix} \tag{2.15}
$$

This covariance matrix is not diagonal, because the axes are correlated (due to the shared component ε_t^1 in the equations defining the relative evidence states), which, when modelling the process, results in an ellipse whose axes are not parallel to the coordinate axes (see the left panel on Figure 2.7). To eliminate the correlation between the axes, we can place them at a more convenient position, by rotating both counterclockwise until the ellipse axes are parallel to the coordinate axes (as demonstrated in the right panel of Figure 2.7).

Figure 2.7: Demonstration of the correlation problem. The left panel shows the original axes, and the right panel shows the rotated axes with no correlation.



In this rotated coordinate system, the probability distribution over the relative evidence states can be characterised by a bivariate normal distribution of two independent variables, allowing us to substantially simplify the subsequent calculations by relying on the normality of the joint distribution of the two relative evidence states. The rotated axes, x' and y' can be characterised (Anton & Rorres, 2014) as

$$\begin{aligned} x' &= xcos(\alpha) - ysin(\alpha) \\ y' &= xsin(\alpha) + ycos(\alpha). \end{aligned} \tag{2.16}$$

In Figure 2.7, the angle of rotation is $\alpha = 45°$. Expressed in terms of our original relative evidence states, this 45° degree anti-clockwise rotation means that the new axes can be defined as

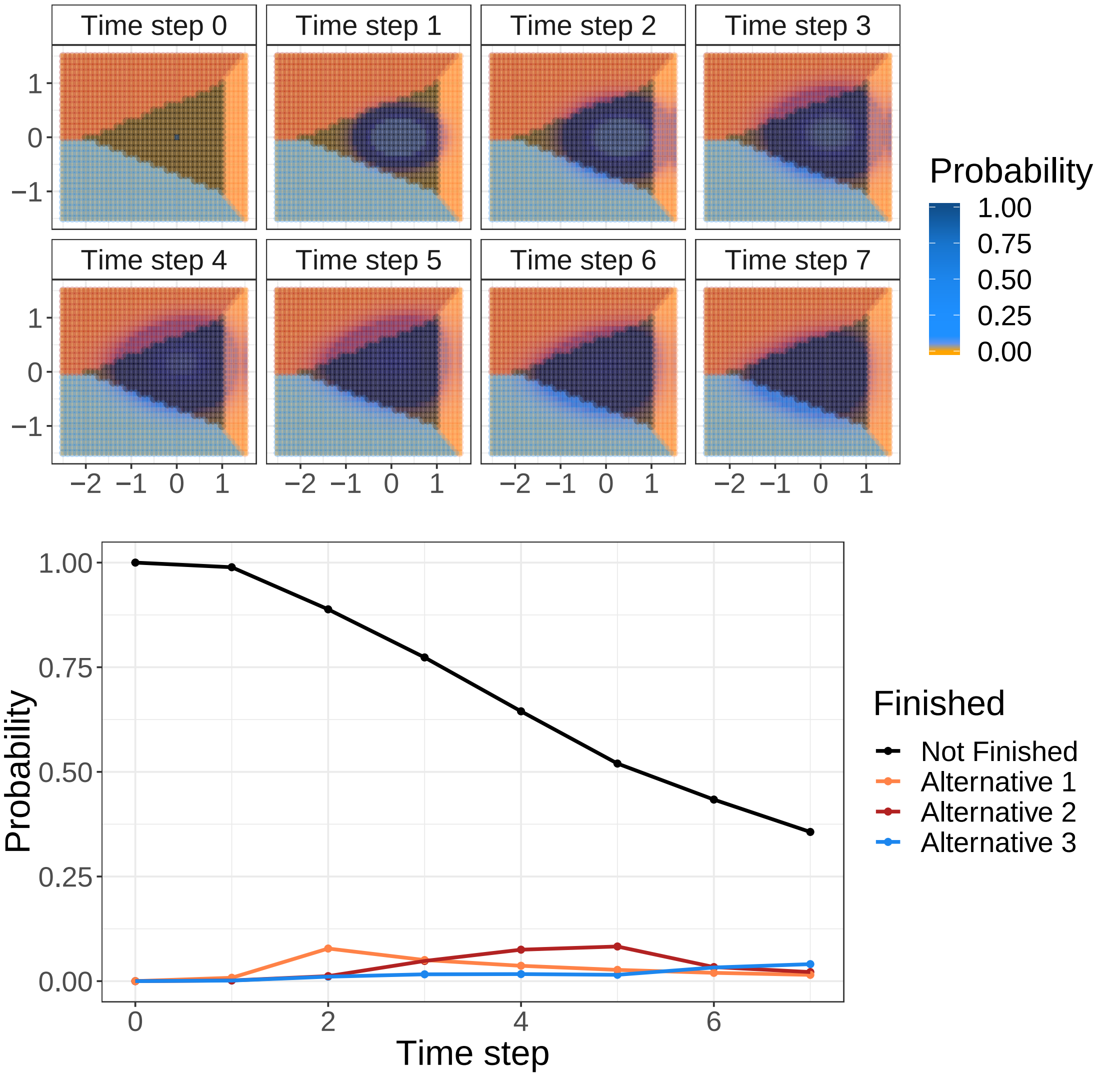
$$\begin{aligned} x' &= E_1 - \frac{E_2 + E_3}{2} \\ y' &= \frac{E_2 - E_3}{2}. \end{aligned} \tag{2.17}$$

As the right panel in Figure 2.7 demonstrates, this results in an ellipse whose axes are now parallel to the coordinate axes. In this transformed coordinate system, we can characterise the decision process as a multivariate normal distribution, whose movement along the relative evidence bins is governed by the model parameters. To save computation time, the possible evidence states were approximated by bins (the x evidence space spanned from -2.5 to 1.5 with spacing 0.1 in 41 bins, whereas the y evidence space spanned from -1.5 to 1.5 with spacing 0.1 in 31 bins, resulting in $31 \cdot 41 = 1271$ bins overall). Our aim was to calculate the probability distribution of the diffusion process over this grid for each time period in the decision process, which allowed us to directly derive the probability that an item is chosen at any given time point, by summing up the probability mass over the evidence bins in which the item was chosen (see the finished evidence states in Figure 2.6).

Figure 2.8 illustrates the diffusion process. On each trial, at time step 0, the process starts from the central bin defined by $x = 0$ and $y = 0$, with probability density 1, corresponding to no evidence for any of the three alternatives.

As shown in the upper panel of Figure 2.8, at each subsequent time step, we can model the movement and diffusion of this probability mass over our evidence grid. In this example, in the first two time steps, Alternative 1 is fixated, and thus the overall probability mass starts moving towards the east (approaching the relative evidence bins that correspond to Alternative 1 being chosen). The movement of the probability mass throughout the choice process is governed by the model parameters. At each time step, we can derive the choice probability that either of alternatives are chosen by summing up the probability over the relevant evidence states. The corresponding choice probabilities are depicted in the lower panel of Figure 2.8. Below we explain in more technical detail how the choice probabilities were derived.

Figure 2.8: Illustration of the diffusion process over 7 time steps with attention pattern $1, 1, 2, 2, 2, 3, 3, 3$; $\sigma = 0.3$; $\theta = 0.67$; $d = 1$, $r = 0.5, 0.5, 0.5$



From a computational perspective, we derive these probabilities with the help of a transition matrix. In each time period, we calculate the probability distribution over the grid by multiplying the previous time period's probability mass with our transition matrix. This matrix reflects the dynamics of the diffusion process, and along with the drift rate, determine the resulting choice probabilities at each time step.

The transition matrix characterizes the diffusion of the probability density with σ as the only input to a bivariate normal distribution with mean $\mu_x = 0$ and $\mu_y = 0$. Specifically, this is done by taking each of the 1271 bins in the grid as the centre of a bivariate normal distribution with standard deviation σ, and calculating the exact probability mass moving from one bin into each of the other 1271 bins in the grid. Calculating this probability vector for all our bins in the grid gives us a $1271 \cdot 1271$ transition matrix, where each entry is the transition probability from one bin to another. Starting with a probability mass entirely concentrated in the bin defined as $x = 0$ and $y = 0$ at time step 0 (corresponding to no evidence for any of the alternatives), we can recursively multiply the resulting probability mass in each time step with this transition matrix. In absence of the second component, this process leads to a diffusion of the probability density around our starting point $x = 0$ and $y = 0$, corresponding to a diffusion process with no drift rate (also called a random walk process).

The drift rate is defined by parameters θ, r and d (described in Equations 2.1–2.3 in Section 2.2), and it governs the direction of the movement of the probability mass. This drift rate can be incorporated in the transition matrix by shifting the mean of the bivariate normal distributions. Because the drift rate varies depending on the attention pattern (through θ), three transition matrices can be constructed, depending on which item is fixated. Taken together, through the model parameters, the transition matrix and the drift rate determine the movement of the bivariate distribution in our grid.

As explained above, we can then derive the probability of choosing either of the options in at any given time period by summing up the probability mass entering the bins corresponding to the option's finished evidence states in that time step. Before moving to the next time period, all the probability mass over the finished evidence states are set to 0, allowing us to derive the probability that an option is chosen exactly in the next time period. To ensure that this estimation method is completely free of stochasticity, we used a deterministic fixation pattern, based on which item was fixated first in each 200 ms bin.

Once we were able to calculate the choice probabilities using this method, we again started off with a grid search to find the best fitting parameters for each participant and subjective value transformation rule. We used the following parameter grid: $\theta = \{0.33, 0.67, 1\}$, $\sigma = \{0.1, 0.55, 1\}$, d=$\{0.1, 0.55, 1\}$ in running the Nelder-Mead optimization algorithm. We used a smaller grid, because there is no stochasticity in this method, but it is also computationally more intensive to calculate (calculating the log-likelihood for one participant takes 13 times longer compared to the simulation approach). For the same reason, we only ran the Nelder-Mead process once more after the grid search, with a maximum iteration number of 100.

2.7.3 Results

For both empirical methods (simulations and probability distribution approach), the model fitting process resulted in 200 log-likelihoods, one for each value transformation rule–participant pair. We wished to determine which value transformation rule produces the highest log-likelihood for each participant and estimation method. For each participant, the four subjective value variants of the aDDM have the same number of parameters, and are calculated on the same set of data, therefore the log-likelihoods are directly comparable. Table 2.2 shows the distribution of participants based on their respective best fitting value transformation rules, demonstrating that the two estimation approaches have yielded very similar results.

Table 2.2: Best fitting value transformation rule for each participant by estimation method.

Approach	Global Max	Local Max	Rank	Range
Simulations	29 (58%)	11 (22%)	4 (8%)	6 (12%)
Probability Distribution	28 (56%)	9 (18%)	8 (16%)	5 (10%)

For the majority of participants (56–58%), the global maximum rule provided the best fit, regardless of the estimation approach. This is followed by the local maximum rule, which provided the best fit for 18–22% of the participants. The range rule proved to be the best fitting model for only about 10–12% of participants. Interestingly, when using the probability distribution approach, the rank rule provided the best fit for 16% of participants, whereas the same number for the simulations approach is only 8%.

Table 2.3 shows the median estimated parameter values by value transformation rule and estimation method. The parameter estimates do not vary considerably across value transformation rules or estimation methods. For the list of best fitting parameter values for each participant by value transformation rule and estimation method, see Appendices A.1–A.2.

Table 2.3: Median best fitting parameter values by value transformation rule and estimation method.

	Simulations Approach			Probability Approach		
	θ	σ	d	θ	σ	d
Global Max	0.68	0.2	0.31	0.7	0.19	0.29
Local Max	0.67	0.2	0.25	0.68	0.19	0.24
Range	0.54	0.2	0.15	0.56	0.2	0.14
Rank	0.51	0.21	0.16	0.53	0.2	0.14

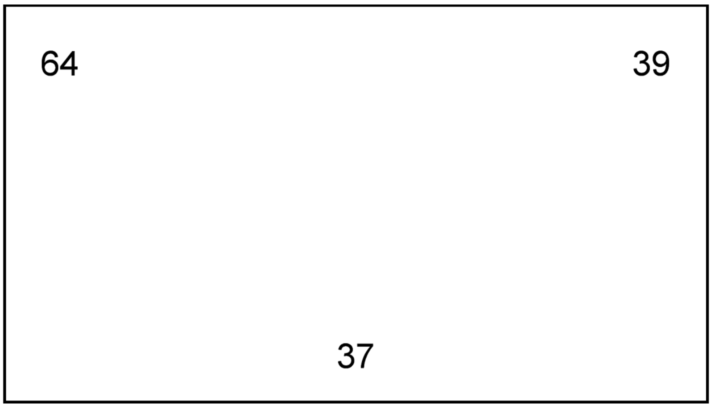
Overall, these results point to a cognitive mechanism where subjective valuation largely depends on the absolute value of the options. The almost identical results from the two estimation approaches underline the robustness of the findings, and alleviates concerns about the reliability of the simulations method. However, it can be argued that the preference ratings participants gave in the rating task are already context-dependent, since the movies are inevitably evaluated with respect to each other. If this was the case, it would bias our findings, which is why we decided to investigate the same question with a perceptual version of Experiment 1.

2.8 Experiment 2

In Experiment 2, we still relied on the choice set manipulations to test the explanatory power of the four value transformation rules, but we used an alternative empirical approach and different stimuli domain. We assessed the explanatory power of the four value transformation rules by their ability to qualitatively predict how the choice set manipulations affect the choice proportions, using choice proportion data from Experiment 1 and a perceptual version of the same task (Experiment 2).

Experiment 2 had several rationales. First and foremost, we wished to alleviate concerns about the internal validity of the experimental design used in Experiment 1. Second, comparing the results from a further comparison of choice probabilities with the results from a model fitting approach allows us to gain a deeper understanding of the relative explanatory power of each value transformation rule. While we have some choice data on the seven relevant choice sets from Experiment 1, due to the error in the experimental procedure, it is not enough to conduct a test with appropriate statistical power. Finally, since the movie stimuli are somewhat complex in the sense that there are a range of factors we could not control for (e.g. visual saliency, mnemonic processes the stimuli can induce), repeating the experiment with simpler and more neutral stimuli will inform us about the validity and generalisability of our results.

Figure 2.9: Example trial from Experiment 2.



2.8.1 Method

We recruited 130 participants through the University of Warwick's Research Participation System. On each trial, participants were presented with a trinary set of rapidly changing sequences of numbers (taken from Tsetsos et al., 2012) and were asked to choose the sequence with the highest average of numbers by pressing the corresponding keyboard key (left, down, right arrow). The numbers were updated every 210–250 ms. To familiarise participants with the task, the experiment began with three practice trials, where feedback was given on which sequence they chose. After the practice trials, there were 140 trials to be completed in two blocks (20 trials per choice set). Participants could take a break for as long as they wished between the two blocks, and they were paid £5.00 for taking part in the experiment. Figure 2.9 shows one trial from this experiment. The experiment on average took about 15 minutes to complete. Ethical approval was obtained from the Department of Psychology, University of Warwick.

2.8.2 Stimuli

Numbers for each presented sequence were derived from a truncated normal distribution with $\sigma = 18$, and a mean that was calculated from a monotonic transformation of the ratings from Experiment 1. Specifically, we multiplied each number by ten, and further added ten to them. For example, a trial with values 1,2,4 in Experiment 1 corresponded to a trial with means 20, 30, 50 in Experiment 2. The distribution was truncated to only include numbers between 0 and 99. We chose to work with the two-digit equivalents of the values from Experiment 1 to increase the range of numbers that could be sampled, and the constant was added to shift the lower end of the value distributions away from 0. When a number below 10 was sampled, it was displayed as a two-digit number with 0 as the first digit to prevent any differences in saliency between one- and two-digit numbers.

2.8.3 Results

To conduct a further test of the explanatory power of the four value transformation rules, we compared the empirical choice proportions (from Experiment 1 and 2) with the choice set manipulation predictions of each value transformation rule. The choice proportion predictions for each value transformation rule were derived from simulating out each trial 10,000 times in the DDM framework. These simulations focused only on the eventually chosen item, and did not take reaction times into account.

In the DDM, there are two parameters, σ, the noise, and d, the speed of integration. While our main interest, the qualitative patterns of predicted choice proportions are largely insensitive to the exact parameter values, we still needed to determine the exact parameter values to use for generating predictions for each value transformation rule. To derive these (group-level) parameter sets, we again used a Nelder-Mead optimization algorithm minimising the squared difference between the simulated and empirical choice proportions for each value transformation rule. The empirical choice proportions were based on the results from Experiment 2 as opposed to Experiment 1, due to the much larger sample size for the relevant choice sets. We ran the optimization algorithm with a maximum iteration number of 100 three times, first with starting parameters

$\sigma = 0.1, d = 0.1$, and then using the resulting parameters from the first and second run as starting parameters for the second and third optimization process, respectively. Table 2.4 shows the best fitting parameter pairs for each value transformation rule.

Table 2.4: Best fitting group-level parameter values from the grid search for each value transformation rule.

	σ	d
Global Max	0.12	0.08
Local Max	0.17	0.07
Range	0.13	0.07
Rank	0.15	0.05

Figures 2.10–2.13 show the predictions for each rule and the associated empirical choice proportions from both experiments. In these figures, the four panels to the left correspond to the choice proportion predictions from the DDM simulations, while the two panels on the right show the resulting choice proportions from the two experiments and their associated 95% CIs. All model predictions are based on 100,000 simulations. The difference in the precision of the estimates from the two experiments reflects the difference in the amount of choice data available for the relevant choice sets.

Adding a constant. Figure 2.10 shows the effect of adding a constant. As described in section 2.5, the local maximum rule predicts that this value manipulation results in the best option becoming relatively less attractive compared to the other two, whereas none of the other three value transformation rules predict any change in the choice proportions. Unfortunately, the wide confidence intervals around the choice proportions from Experiment 1 do not allow us to see any clear patterns, but the results from Experiment 2 are broadly in line with the predictions from the local maximum rule (albeit the changes are less pronounced than predicted). In summary, the results from adding a constant lend partial support to the local maximum rule, as no other value transformation rule predicts the correct qualitative pattern.

Figure 2.10: Choice set manipulation adding a constant

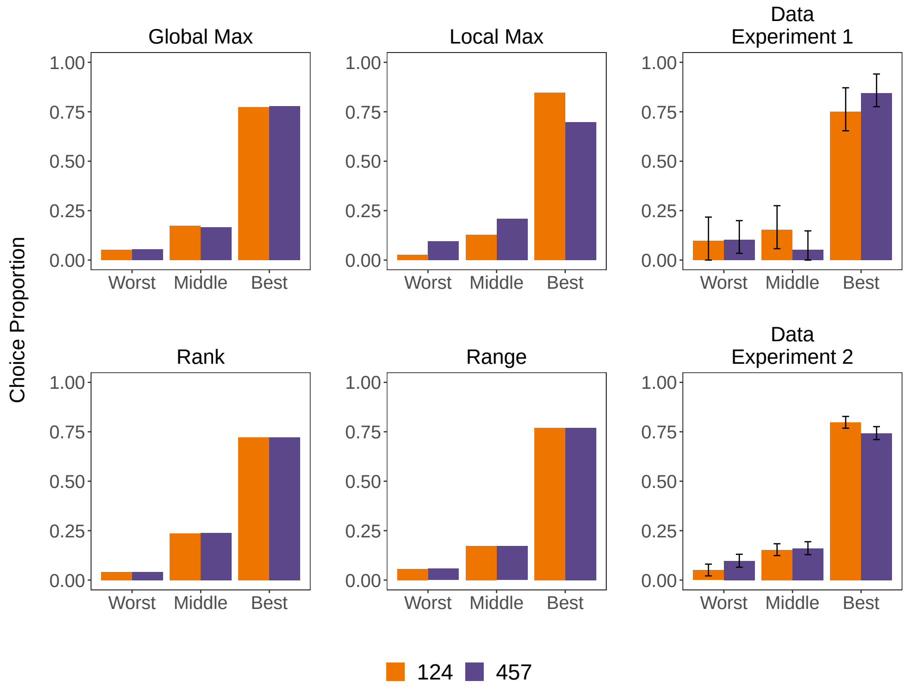
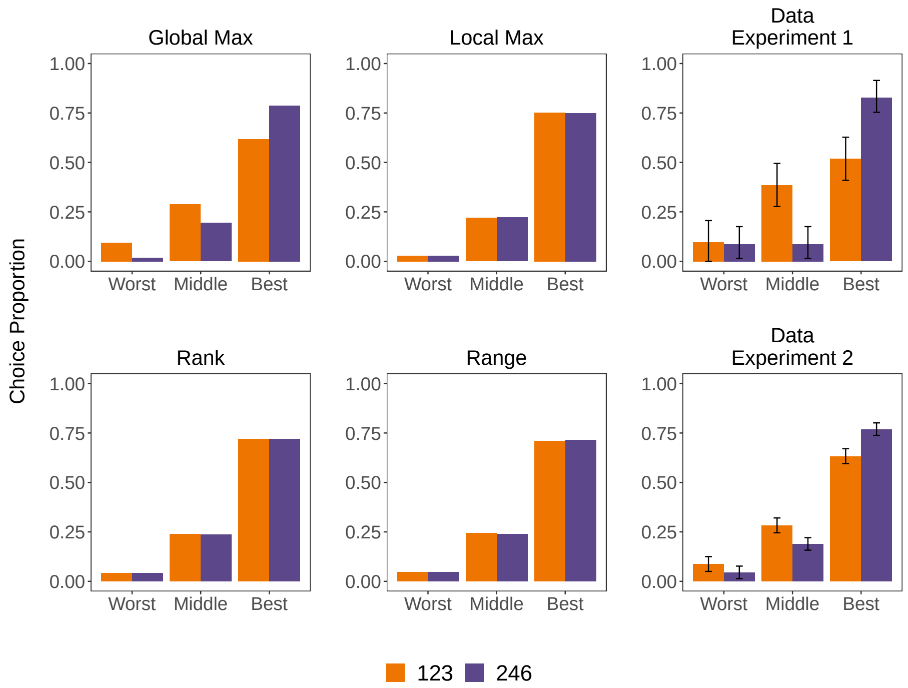


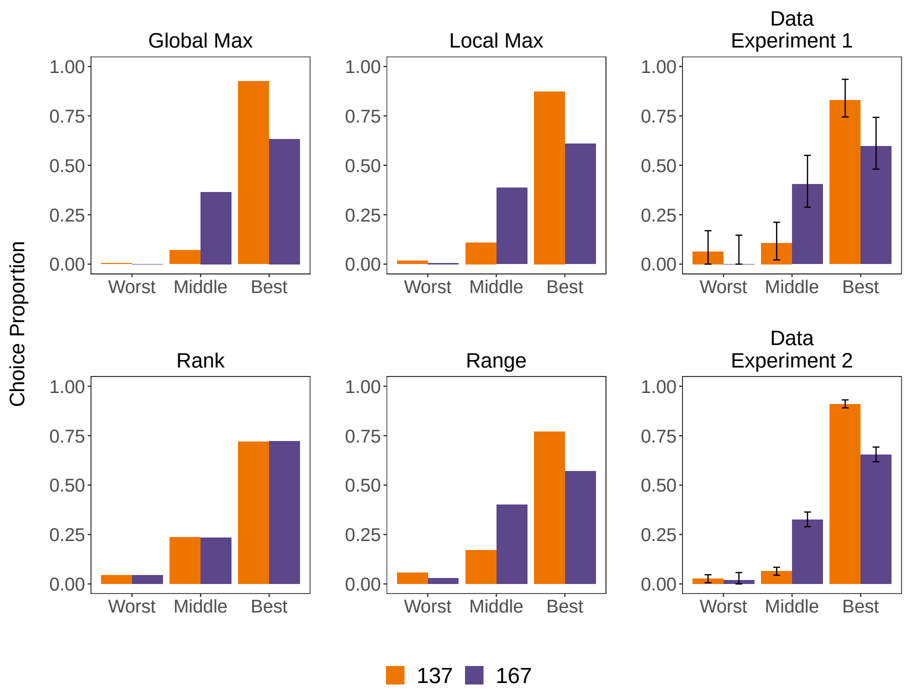
Figure 2.11: Choice set manipulation multiplication by constant



Multiplying by a constant. Figure 2.11 shows the effect of multiplying each value by a constant. The global maximum rule predicts that this value manipulation results in the best option becoming relatively more attractive compared to the other two, whereas none of the other three value transformation rules predict any change in the choice proportions. The results from both experiments are in line with the prediction that the best option will increase its choice proportion share at the expense of the other two, strongly supporting the global maximum rule.

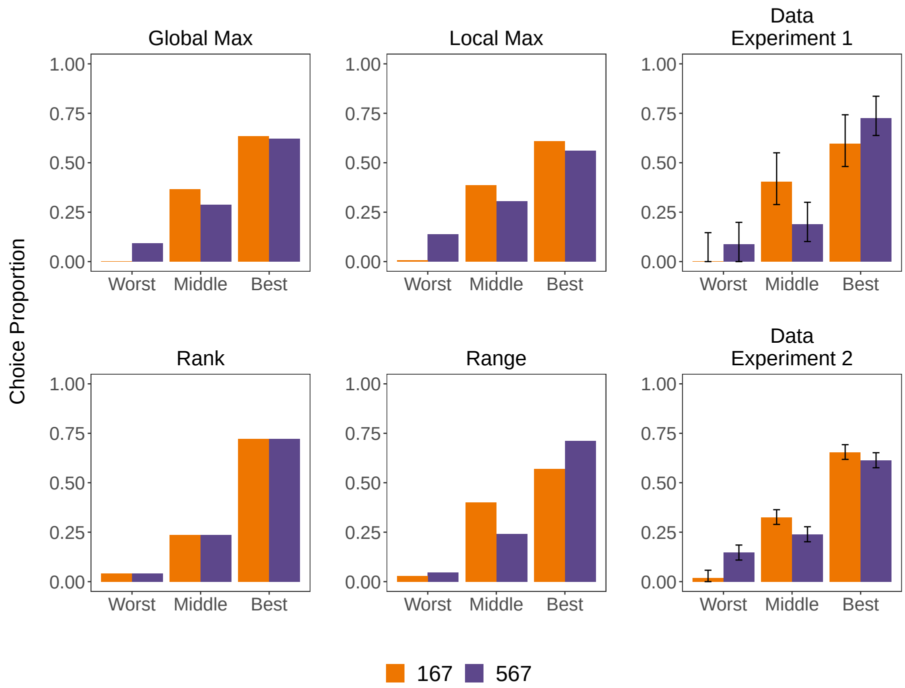
Distant versus close second. Figure 2.12 shows the results from the close versus distant second choice set manipulation. Only the rank rule does not predict any change in the choice proportions, while the other three rules predict that the middle option becomes relatively more attractive. Results from both experiments support the latter prediction. Therefore, results from the distant versus close second choice manipulation provides strong evidence against the rank value transformation rule.

Figure 2.12: Choice set manipulation distant versus close middle



Distant versus close third. Finally, Figure 2.13 shows the results from the close versus distant third choice set manipulation. The predictions of the value transformation rules differ significantly, the two maximum rules predicting that the worst option will benefit at the expense of the other two (albeit the extent of this change slightly differs between the two rules), and the range rule predicting that both the worst and best options will steal from the middle item's share. The rank rule does not predict any change. While the results from Experiment 2 unequivocally support the predictions of the two maximum rules, the results form Experiment 1 are more mixed (although we can detect a clear decrease in the choice proportion of the middle item). Therefore, results from the close versus distant third choice manipulation offer support for the two maximum rules.

Figure 2.13: Choice set manipulation distant versus close third



To summarize, the results from these qualitative comparisons suggest that the two maximum normalisation rules had the highest explanatory power. More specifically, albeit to varying degrees, both maximum rules predicted the effect of three out of the four choice set manipulations, the range rule successfully predicted one, and the rank rule did not predict the effect of any of the choice set manipulations.

These results can be contrasted with the results from the model fitting approach, which indicated that the global maximum rule has the highest explanatory power by far. In line with this, multiplication by a constant, which was only correctly predicted by the global maximum rule, resulted in the largest difference between choice proportions in Experiment 1. However, it is also clear from the results of the qualitative comparison that the global maximum rule cannot alone explain choice behaviour.

In addition, interestingly, even though the two experiments involved fundamentally different stimuli (complex versus perceptual), the effects of the choice set manipulations on the choice proportions turned out to be rather similar. Unfortunately, a strict comparison of the two datasets is hindered by the relatively small sample size of Experiment 1. Nevertheless, the results lend some support to the idea that there exists a common valuation mechanism across a range of stimuli domains.

2.9 General Discussion

In this research project, our aim was to conduct a comprehensive evaluation of the relative explanatory power of four different forms subjective valuation. To compare four value transformation rules, each of which captures a different subjective valuation mechanism, we used a popular cognitive model, the DDM and its extension, the aDDM, to derive predictions about choice behaviour in two choice experiments with trinary choice sets, one with complex stimuli (movie posters), and another one with a perceptual task (rapidly updating sequences of numbers).

We tested the relative explanatory power of these value transformation rules in three separate tests. In the first two tests, we used a model fitting approach (based on simulations in the first test, and approximated choice probabilities in the second test), using choice and eye-tracking data from Experiment 1. The results showed that the global maximum rule was the best fitting subjective transformation rule for the majority of participants, followed by the local maximum rule. In addition, the results from the two tests were remarkably similar, alleviating concerns about employing a simulations-based estimation approach in the context of a stochastic model.

In the third test, we compared the qualitative patterns of choice proportions using data from Experiment 2 and 1. The results from the two comparison methods were broadly in line, both supporting the view that while subjective valuation mostly reflects the absolute magnitude of the options under consideration, to a lesser extent, it is also affected by the relative value of the options (within the local context).

These findings are consistent with insights from research investigating the neural correlates of value-based decision making. Results from numerous neuroeconomic studies strongly support the view that the orbifrontal cortex (OFC) is the brain region where subjective value encoding takes place in economic choices (for a review see Padoa-Schioppa & Conen, 2017). More importantly, it has been suggested that there is a group of neurons in the OFC with two fundamental properties that are likely to be directly responsible for simultaneous absolute and relative value sensitivity in economic valuation.

For example, Padoa-Schioppa and Assad (2008) have shown that such neurons exhibit menu invariance, meaning that the value assigned to each option under consideration is independent from the value of the rest of the available alternatives, reflecting the absolute value of the option (global maximum in our experiments). Menu invariance gives rise to preference transitivity, ensuring that preferences are stable across the wide range of contexts the decision maker might encounter. In addition, there is ample evidence that neuronal firing rates adapt to the range of available values (e.g., Padoa-Schioppa, 2009; Louie et al., 2013). Such adaptation is widespread in sensory systems, and is a natural consequence of biophysical constraints. Interestingly, although the most commonly proposed form of adaptation in the neuroeconomic literature is range adaptation (e.g., Soltani et al., 2012), our data suggests that normalising by the maximum on the current trials fares better than a range normalisation approach at predicting choice.

There exist other ways to model context sensitivity within the sequential sampling framework. Another approach to investigate context dependency could have been to directly model changes in the drift rate. This could be done within an experiment where the range of values encountered are manipulated block by block (e.g., as done in Mullett & Tunney, 2013), and the model fitting approach can evaluate the explanatory power of various forms of context dependency.

In addition, we could have investigated mixtures of models, for example, a hybrid model, where the subjective value is partly affected by the absolute value and partly affected by the local maximum transformation. Such investigations would have required estimation an additional, mixing parameter, which determines the degree to which each rule affects the subjective value.

Alternatively, instead of changing the input values of the accumulation equation, we could have focused on how these values are integrated and incorporated into the accumulation process (described in Equations 2.4–2.6). This is what Teodorescu, Moran, and Usher (2016) did in a somewhat similar investigation to ours. Specifically, they contrasted relative and absolute evidence processing in a sequential sampling framework by comparing an independent race model (capturing absolute value processing, where the input is the absolute value of the options), with a DDM model (where differences of input values govern the accumulation process).

In their experiment, participants were instructed to choose the brighter out of two, fluctuating grey patches, with a fixed mean brightness. They focused on two manipulations: in an additive-boost condition, they added the same constant to both means, preserving the difference between the two mean brightness, whereas in the multiplicative-boost condition, they multiplied both means by the same constant, preserving the ratio of the two means. These conditions are direct equivalents to our add a constant, and multiply by a constant choice set manipulations.

Interestingly, they found that no "pure" (either entirely absolute or relative) accumulation model could account for the data, and thus they propose two distinct types of models that can account for this pattern: a DDM model where the noise in the process is a function of the intensity of the inputs, and a leaky competing accumulator model (LCA; Usher & McClelland, 2001), where simultaneous absolute and relative value sensitivity is a result of lateral inhibition.

While our approach is different from that of Teodorescu et al. (2016), our results are similar, as they both suggest that subjective valuation is sensitive to both the absolute and relative magnitudes of objective values. This dovetails with findings from neuroeconomic research that suggests that OFC neurons exhibit both menu invariance and range adaptation. Taken together, these results point to a sequential sampling model with some form of hybrid value transformation rule. As an ever increasing amount of research focuses on understanding the neural basis of economic decision making, insights from this field will no doubt inform and greatly advance the explanatory power of cognitive models of choice in the future.

Chapter 3

Testing the attraction effect with naturalistic stimuli

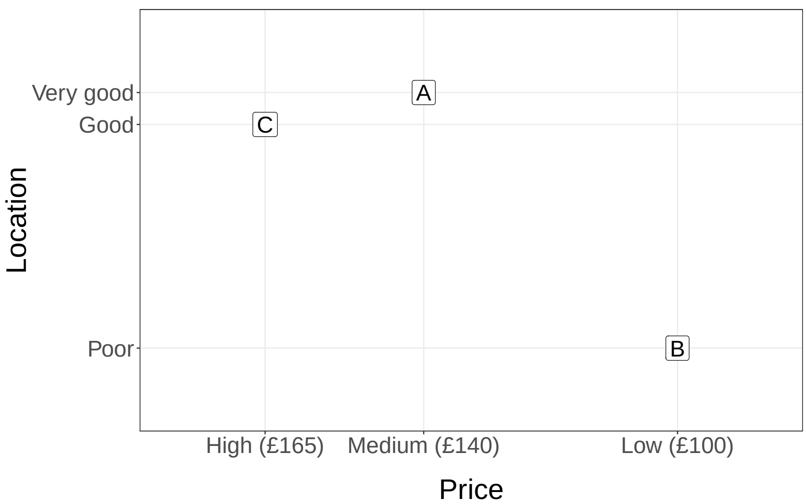
3.1 Introduction

3.1.1 The attraction effect

Suppose you are looking to book a hotel for a weekend city break. You first encounter two hotel options: Hotel A with an excellent central location for £140, and Hotel B with a much less central location for £100. At this point, you are not quite sure whether the better location of Hotel A is worth the extra money. Then you notice that there is a third option, Hotel C, with a slightly less central location than Hotel A, for £165. Which one would you choose?

A large body of decision-making research suggests that the presence of Hotel C will make it more likely that you will go for Hotel A. Figure 3.1 illustrates this choice situation, placing the three hotels in a price-by-location attribute space. This phenomenon is called the *attraction effect* (also known as decoy effect or asymmetric dominance effect), and it has been researched extensively since the first time it was demonstrated by Huber, Payne, and Puto (1982).

Figure 3.1: The attraction effect: Hotels A, B and C differ on price and location. Assuming that the decision maker is indifferent between Hotels A and B, the introduction of Hotel C that is inferior to Hotel A on both attribute dimensions makes it more likely that Hotel A will be chosen.



This effect poses a challenge to all choice models that rely on the assumption that preferences can be represented on a cardinal scale in the form of utilities (this is also called simple scalability, one of the consequences of Luce's Choice Axiom; Luce, 1959). Such models must satisfy two axioms. The first is called order independence (its stricter version is known as independence from irrelevant alternatives), which requires that the preference ranking of two options should not be affected by adding new options to the choice set. The second axiom is called regularity, and it states that by extending the choice set, an option's choice probability cannot increase (Luce, 1977; Tversky, 1972).

The attraction effect clearly violates these axioms, as it shows that the addition on an inferior decoy option increases the preference for the target over the competitor. This means that the preference for one option cannot be represented by a single internal magnitude that is invariant to the other options' value in the choice set. In fact, the presence of the effect supports a choice model where the preference for a choice option is strongly affected by the range of available options through the comparison process.

The attraction effect has been studied extensively over the past few decades using various experimental populations and conditions, as well as stimuli types. This research had shown that compared to young adults, older adults are less susceptible to this decision bias (Kim & Hasher, 2005), and that children as young as 5-year-olds already exhibit this effect (Zhen & Yu, 2016).

Several studies have investigated whether such choice biases are also present in animal decision making. This line of research is very important as it has the potential to deliver valuable insights into the evolutionary origins of the attraction effect and related choice biases, informing attempts to model human decision making. Unfortunately, results from this area so far seem fairly inconclusive. While decoy effects have been observed in various animals from brainless amoebas (Latty & Beekman, 2011), to honeybees, grey jays (Shafir, Waite, & Smith, 2002) and rhesus macaques (Parrish, Evans, & Beran, 2015), a similar number of studies report conflicting results (in rhesus macaques, Parrish, Afrifa, & Beran, 2018; capuchin monkeys, Cohen & Santos, 2017; ant colonies, Edwards & Pratt, 2009; and hummingbirds, Bateson, Healy, & Hurly, 2002).

It is possible that there is no single, coherent explanatory framework that could accommodate these results, as the aforementioned studies have utilized a wide range of experimental tasks and used very different subject populations, thus they might only be superficially similar. In any case, these inconclusive results highlight the attraction effect's sensitivity to the experimental task and population, and stresses the importance of exploring the boundary conditions of the effect.

In humans, the strength of the effect has been probed under various experimental conditions. Research in this area showed that the attraction effect is more pronounced under time pressure (Pettibone, 2012), and less pronounced with undesirable choice options (Malkoc, Hedgcock, & Hoeffler, 2013), and feedback after choice (Ahn, Kim, & Ha, 2015). Mao and Oppewal (2012) found that the strength of the effect is also influenced by individual differences: in their experiment, participants who relied more on intuitive reasoning showed a stronger attraction effect.

The role of individual heterogeneity has also been the focus of recent research that attempted to accommodate the three most well-known decision biases (the attraction, similarity, and compromise effect, also known as the Big Three) in a single modelling framework. Previous investigations have found that while the size of the attraction and compromise effect are positively correlated, both are negatively correlated with the size of the similarity effect (Liew, Howe, & Little, 2016; Berkowitsch, Scheibehenne, & Rieskamp, 2014). A recent study by Cataldo and Cohen (2018) offers a simple, plausible explanation for this finding.

In this study, they manipulated the presentation format of the choice options (such that the presentation format facilitated either an alternative-wise or an attribute-wise comparison process), and found that the attraction and compromise effects were stronger in the condition where participants were encouraged to use an attribute-wise comparison process, whereas the similarity effect was stronger in the alternative-wise condition. This finding suggests that any modelling attempt that aims to offer an explanation for all three decision biases must take into account how individual heterogeneity affects the comparison strategy.

Several explanations have been put forward to explain the attraction effect. Simonson (1989) proposed that the decoy makes it easier to justify the subsequent choice of the target. Huber et al. suggested that the presence of the decoy changes the reference level against which the target and the competitor are evaluated, making the target seem relatively more attractive. With the added assumption of loss aversion, prospect theory also relies on reference dependence to explain the attraction effect (Sivakumar, 2016). Another explanation that had been proposed is trade-off aversion (e.g., Hedgcock & Rao, 2009). According to this theory, the attraction effect stems from the decision maker's inherent desire to avoid negative emotion. This is because the presence of the decoy reduces the cognitive cost and any potential negative affect generated during the evaluation and subsequent comparison of two distinctly different choice options.

The latest modelling efforts aiming to explain the attraction effect almost exclusively focused on sequential sampling models of choice. The attentional drift diffusion model (aDDM) described in Chapter 2 is one such model. As mentioned, these models view the choice as a series of noisy evidence accumulation steps in favour of each choice option, where the option that first reaches a given threshold will be chosen. Prominent examples of these kind of models are multi-alternative decision field theory (MDFT; Roe, Busemeyer, & Townsend, 2001), multi-attribute decision by sampling (MDbS; Noguchi & Stewart, 2014), the leaky competing accumulator model (LCA; Usher & McClelland, 2004), the multi-attribute linear ballistic accumulator model (MLBA; Trueblood, Brown, & Heathcote, 2014), and the associative accumulation model (AAM; Bhatia, 2013).

These sequential sampling models became very popular in decision making research as they can explain the attraction effect along with other well-documented decision biases in a single modelling framework. While the models listed above share the same underlying principle, they rely on different mechanisms to produce the attraction effect. Specifically, the MDFT sees the attraction effect as a result of the inhibitory links between alternatives, the LCA relies on loss aversion and inhibition, the MDbS assumes that the presence of the decoy changes the distribution of the evidence samples, the MLBA uses the assumption of attention weights that are inversely proportional to the discriminability of two options, while the AAM produces the attraction effect through the increased accessibility of the attribute dimension on which the target option is the strongest.

Considerable research has focused on the neural mechanism underlying the attraction effect. These studies have typically used fMRI to measure neural activity during choice, and showed that neural activation in certain brain regions depends on the relative value of the option under consideration in the current choice context (e.g., Mohr, Heekeren, & Rieskamp, 2017; Chung et al., 2017; Gluth, Hotaling, & Rieskamp, 2017). These studies also lend support to models that utilize a sequential sampling principle, as neural activity during choice in brain regions that are involved in decision making (specifically, in certain prefrontal and parietal cortical areas) often resemble an accumulation process (Busemeyer, Gluth, Rieskamp, & Turner, 2019). Most importantly, results from studies using neural recordings support the view that value is constructed and therefore is intrinsically dependent on the choice context.

3.1.2 The real-world relevance of the attraction effect

The real-world relevance of the attraction effect has recently become a contentious issue in the decision-making literature. While the effect has previously been demonstrated over several choice domains (e.g., consumer products, Doyle, O'Connor, Reynolds, & Bottomley, 1999; medical decisions, Schwartz &

Chapman, 1999; mate choice, Sedikides, Ariely, & Olsen, 1999; hiring decisions, Highhouse, 1996; political choice, Sue O'Curry & Pitts, 1995; work-family benefits, Reb, Li, & Bagger, 2018; intertemporal choice, Gluth et al., 2017; perceptual decisions, Trueblood, Brown, Heathcote, & Busemeyer, 2013), some claim that the attraction effect is much less prevalent in natural contexts than previously thought.

In their study, Frederick, Lee, and Baskin (2014) present a thorough investigation of the boundary conditions of the attraction effect based on 38 experiments with various stimuli types. These stimuli types include choice options with numerically represented attributes as well as complex, real-world stimuli (e.g., fruits, bottled water, apartments, etc.), and in some of these experiments participants could even sample the choice options (e.g. squash, mints, popcorn). The overall conclusion of this study is that while the presence of the decoy seems to affect decisions when the option attributes are represented numerically, it is absent in experiments with more complex, naturalistic stimuli. In light of these results, Frederick et al. posited that the psychological processes underlying decisions that involve options with numeric attributes are fundamentally different from those employed in decisions where the stimuli has a naturalistic representation. This conclusion was also supported by Yang and Lynn (2014), who reported difficulties replicating the attraction effect when the stimuli were pictorial, as opposed to when attributes were presented numerically.

These two studies sparked considerable interest amongst decision making researchers, and led to the re-examination of the boundary conditions of the attraction effect. Huber, Payne, and Puto (2014) discussed five critical conditions that can inhibit the attraction effect, and argued that many of these are present in the experiments reported by Frederick et al. and Yang and Lynn, which can explain their failure to observe the effect. These are the following: (1) strong prior preferences over the target and competitor, (2) inability to identify the inferiority of the decoy, (3) heterogeneity in prior preferences over the target and competitor, (4) an undesirable decoy, and (5) a too desirable decoy. Simonson (2014) further stressed the importance of the detection of the dominance relationship in observing the attraction effect, and also pointed out several other smaller, specific methodological shortcomings of the studies by Frederick et al. and Yang and Lynn.

While these reactions have seemingly ended the debate, we believe that the literature is still lacking a conclusive answer regarding the presence of the attraction effect in choices with non-numeric option attribute presentations. A rigorous investigation of this question would not only inform us about the real-world relevance of the attraction effect, but would also shed light on the commonalities between the cognitive mechanisms underlying choices that involve options with numeric and naturalistic attributes.

In light of the conflicting views about the importance of this effect, this study is an attempt to replicate the attraction effect using naturalistic stimuli with non-numeric attributes, whilst accounting for the major concerns raised in connection with Frederick et al. and Yang and Lynn's studies. Specifically, our novel experimental design ensures that decision makers are indifferent between the target and competitor and that the inferiority of the decoy is clearly identified. In addition, to increase the statistical power of our test, we used a within-subjects design to identify the effect as well as both A, B, A' and B, A, B' triplet pairs, where X' is the dominated option (as suggested by Huber et al., 2014).

3.1.3 Overview of Experiment 1 and 2

When selecting the type of naturalistic stimuli for testing the attraction effect, our decision was guided by a few simple criteria we considered important for creating target-decoy-competitor triplets. First, we needed stimuli that would be at least of some interest to most people, or at least would be familiar to most of our participants. Second, the stimuli needed to have multiple attributes, which can take a wide variety of values, as most real-world stimuli do. Third, we needed to be able to establish the degree of similarity between pairs of items, which is a non-trivial task in the case of naturalistic stimuli, due to the potentially high number of dimensions and their incommensurability. In addition, we also needed a large number of items to create enough triplets for a powerful within-subjects design.

We decided to use movie posters as stimuli, as it satisfied most of our criteria. Movies are an integral part of popular culture in Western societies, thus we could reasonably expect that most participants will be familiar with well-known movies. They also vary greatly by genre and topic, which offers us a natural way to establish similarity between pairs of movies. In addition, even if we only use reasonably well-known movies, we still have a vast pool of items to create triplets from.

When choosing between naturalistic items such as movies, the decision maker cannot rely on the numerical attributes of the choice options to establish similarity and build a preference ordering (strictly speaking this is not true in a non-experimental context, where people can always at least partly base their decision on the numeric rating of the movie, but this information was not available in our experiment). We expect mnemonic processes to have a very important role in building preference representations of movie items, and while this is inevitable if we are to use complex items that people are familiar with, it can also be potentially problematic. Specifically, we do not exactly know how the retrieval and preference construction process works and the extent to which individual heterogeneity affects it.

However, regardless of how the preference representations are constructed, we expect the comparison process to take place along a few salient attribute dimensions. These dimensions are likely to include the story themes, the genre of the movie, and perhaps the actors and director. Therefore, it is possible to establish a method to create movie pairs that are likely to be perceived different or similar, using all available information on these movies.

In Experiment 1, we used two criteria to establish a measure of similarity between our naturalistic stimuli items. First, we used latent semantic analysis on the text associated with each movie (including various descriptions of the movie, as well as the information on the actors and director). Section 3.1.4 provides a brief overview of what latent semantic analysis entails. Second, we used the overlap in genre categories that movies were assigned to on IMDb as a measure of similarity.

We did not find any evidence for the attraction effect in Experiment 1, but the results also indicated that the perceived similarity between the target and decoy movies was often not strong enough to consider this as a valid test of the attraction effect. Therefore, in Experiment 2, we relied on more detailed genre information from allmovie.com, as well as similarity ratings from an independent experiment to establish similarity. The target-decoy pairs were indeed perceived as more similar in Experiment 2, but we still did not find any evidence for the attraction effect. We speculate that this result arises from differences in the cognitive processes underlying the evaluation of stimuli with numeric attributes versus naturalistic objects.

3.1.4 Establishing complex object similarity: latent semantic analysis

Latent semantic analysis (LSA; also known as latent semantic indexing) is a statistical technique that was developed in the 1990s as a novel method for automatic indexing and retrieval of text documents (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). In LSA, each document in a set is considered as a bag of words (a document here refers to a body of text). Thus the raw data used by the LSA algorithm are huge document-by-word matrices, where entries indicate the presence of a word in a document. Based on these matrices, the LSA algorithm builds a specified number of "latent dimensions", each of which can be represented with a set of words that reflects a common underlying theme in the documents. Then, similarity between any two documents can be established based on the overlap in the latent dimensions that best characterise them. This technique relies on singular value decomposition (SVD), a statistical tool that uses dimensionality reduction to build a simpler, approximate representation of a matrix (Leskovec, Rajaraman, & Ullman, 2014). This tool underlies many real-world applications, from image compression to solving systems of linear equations (e.g., Akritas & Malaschonok, 2004).

Enormous volumes of text data are being generated by users every day, and there is an increasing need to understand the underlying patterns in these texts. One of the most popular techniques used to achieve this is LSA. Accordingly, it has recently gained substantial popularity as a machine learning technique that had been used in various “big data” contexts, from predicting sales performance of movies from online reviews (Yu, Liu, Huang, & An, 2012) to building recommendation systems (e.g. Wu, Wang, & Cheng, 2008), predicting political orientation from Twitter data (Conover, Gonçalves, Ratkiewicz, Flammini, & Menczer, 2011), and detecting online bullying (Bigelow, Edwards, & Edwards, 2016). Our research project largely builds on a study by Bhatia and Stewart (2018), where they successfully used LSA to build multi-attribute representations of choice options and predict subsequent choice behaviour.

To explain how the algorithm works in a nutshell, imagine that we have a set of movies (these are the documents) with the corresponding user-generated text that summarizes each movie. Then, we can create a term-document matrix, where each row corresponds to a unique word found in these reviews, the columns are the movies itself, and the cells represent the number of times each word has appeared in a given movie description. We can call this matrix M_{TxD}, where T is the number of unique words that can be found in our corpus (the rows) and D is the number of movies (the columns). Say we are interested in retrieving k $< \min\{T,D\}$ underlying dimensions from these movie description texts, where k is a measure of the crudity of our simplified representation of these movies: the higher it is, the more detailed our alternative representation of matrix M, but the less efficient it is.

Then, by performing a SVD, we can decompose matrix M the following way:

$$M_{TxD} = U_{Txk}\Sigma_{kxk}V_{kxD} \tag{3.1}$$

where U_{Txk} is a matrix with the loading of each unique word on the underlying k latent dimensions. To put it differently, it contains information about the words that best describe the retrieved latent dimensions, giving interpretable meaning to them. Similarly, V_{kxD} contains the loading of each document on these k latent dimensions. One can consider the V_{kxD} matrix as giving the coordinates of all of the movies in k-dimensional movie space, where U_{Txk} defines each of the dimensions. The k latent dimensions are always retrieved in their order of importance, defined by the share of variance they can explain (thus increasing the number of latent dimensions has a decreasing marginal return in terms of variance explained). Having obtained these matrices, we are ready to 1) describe each movie based on the latent dimensions it has the highest loading on and 2) establish similarity between the movies based on their common topics. Section 3.2.1.2 demonstrates these steps using our movie stimuli.

3.2 Experiment 1

3.2.1 Method

3.2.1.1 Stimuli

As mentioned above, our aim was to select a rich set of movies as stimuli. To this end, we collected data from www.imdb.com on the most popular (given by the number of votes) 800 movies from each of the following genre categories:

comedy, sci-fi, horror, romance, action, thriller, drama, mystery, crime, animation, adventure, fantasy; amounting to 9,600 movies overall. Since most movies appear in multiple categories, if a movie in one category has already been retrieved as part of another category the code continued to the next movie (the retrieval by category happened in the same order as they are listed above). The information we retrieved for each movie were the following: poster image, title, director, actors, number of votes, genre categories, plot keywords and summaries of the movie. Keywords and summaries describe the plot of the movie and are both generated by the users. We used both texts to establish semantic similarity between the movies.

Before performing the LSA, the texts needed to be cleaned. This included removing any text that was not written in English, removing numbers, stop words, punctuation and duplicate words. In addition, we eliminated whitespace from keyword expressions that consisted of multiple words, and thus were treated as one word, to retain their meaning (e.g., expressions like "organized crime"). If, after these transformations, there were no text left for a given movie, it was excluded from the analysis. We had text data with 61,525 unique words for 9,272 movies.

The next step was to create a term-document matrix with each row as a word and each movie as a column. We then used term frequency-inverse document frequency weighting (also known as tf-idf) to measure the importance of each word in each movie text. Assuming that there are $i = 1...T$ rows (words) and $j = 1...D$ columns (movies), and $n_{i,j}$ is the number of times a word occurs in a given document, we applied the following weight to each cell:

$$weight_{i,j} = \frac{n_{i,j}}{\sum_{i=1}^{T} n_{i,j}} \log\left(\frac{D}{\sum_{j=1}^{D} \min\{n_{i,j}, 1\}}\right) \tag{3.2}$$

In other words, for a given word and document, the weight is the product of the 1) share of the word from all the words in that document multiplied with

2) the logarithm of the overall number of documents divided by the number of documents that contain the word.

The underlying principle of this weighting scheme is that the importance of a term is inversely proportional to the number of documents it appears in. For example, "murder" and "kamikaze" are both keywords used to describe the plot of Pearl Harbor, but since "murder" appears in 3,336 movie descriptions, whereas "kamikaze" only appears in 10, the latter is assumed to contain more information about the movie, and therefore will be assigned a higher weight. We performed the latent semantic analysis on this cleaned, weighted corpus.

3.2.1.2 Choice set selection

We decided to use a latent semantic solution with 20 latent dimensions, as we found that this number gives us a sufficiently rich latent dimension space to establish similarity between movies, while keeping computational complexity at a reasonable level. Figure 3.2 shows the 10 first words with the highest absolute loading on these 20 dimensions (this is part of the U_{Txk} matrix from Equation 3.1). Words that have an orange colour load positively on the given dimension, while words in blue load negatively on them. This means that some of the dimensions have a somewhat counterintuitive "reversed definition", and are described by words that are the *least* characteristic of them.

For example, the words that best describe latent dimension number 4 are "slasher", "maniac", "homicidal maniac", which all relate to a serial killer story, so we would expect serial killer movies to load highly on this dimension, whereas movies with no such theme (e.g., romantic comedies) to load negatively on this dimension. Similarly, latent dimension number 5 has high negative loadings for words that relate to interpersonal relationships, such as "marriage", "romance", "married", so we would expect romantic movies to load negatively on this dimension (as the dimension can be best described as "non-romantic"), whereas any movie without such themes (e.g. animated superhero movies, zombie movies) should load positively on this dimension.

Table 3.1 shows the ranking of the 20 latent dimensions for four selected movies. The ranking is based on the absolute value of the movie's loading on a given dimension (this was derived from matrix V_{kxD} from Equation 3.1), such that the latent dimensions are listed in descending order of relevance for each movie. Similarly to Figure 3.2, the colour code shows the direction of the association: the dimension number is coloured orange if the movie loads positively on that given dimension, and blue if it loads negatively.

Based on the information presented in Figure 3.2 and Table 3.1, we can now attempt to describe each of the four movies based on what the algorithm tells us about them. Naturally, dimensions with positive loadings are more helpful if one wishes to use the latent semantic solution to build a lower-level representation of each movie. For example, assuming no prior knowledge about these four movies, the latent semantic analysis solution tells us that Deadpool is a superhero movie with adult themes (such as nudity and violence), Love Actually is a movie set in the real world and is about romantic relationships, Psycho features homicide and is set in the US countryside, while Wolf of Wall Street is a movie that is set in the US with lots of adult scenes (mostly sexual, but there is violence too) that features an investigation.

While even a human might find it hard to accurately describe a large, diverse set of movies with only 20 "themes", the algorithm seems to perform surprisingly well in capturing the most important aspects of these movies, highlighting the efficacy of the LSA algorithm.

Figure 3.2: The ten most important terms (with the highest absolute loading) for each of the 20 latent dimensions. Orange coloured terms load positively on the latent dimensions, whereas blue coloured terms load negatively on the dimensions.

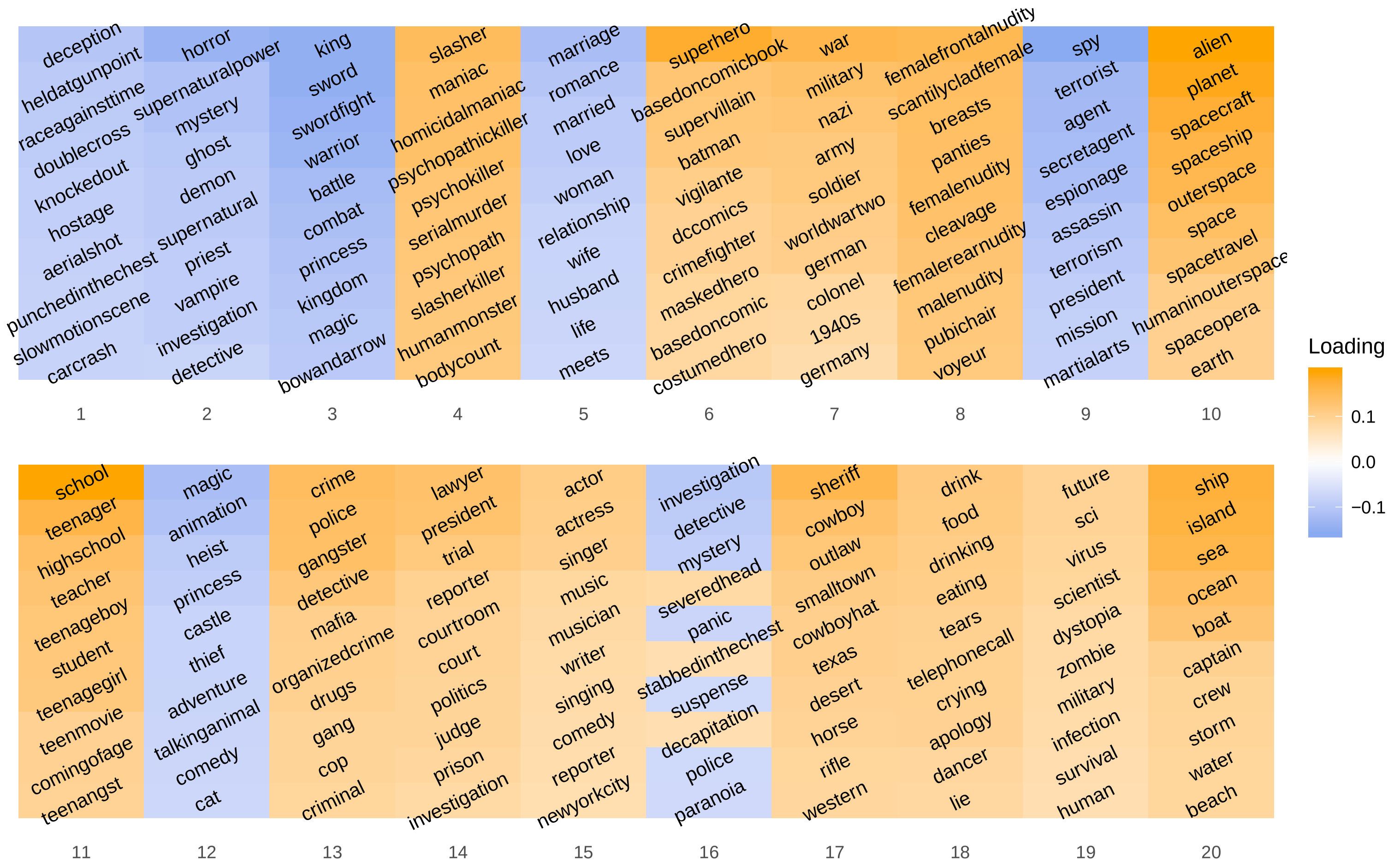


Table 3.1: Ranking of the 20 latent dimensions for each of the four selected movies (based on the movies' absolute loading on each dimension), in descending order of relevance. The movies load positively on orange coloured latent dimensions, and negatively on blue coloured dimensions.

Rank	**Deadpool**	**Love Actually**	**Psycho**	**The Wolf of Wall Street**
1	6	15	4	14
2	16	5	5	8
3	1	8	17	20
4	12	13	2	16
5	7	12	16	19
6	8	14	11	7
7	14	2	12	15
8	10	1	13	1
9	19	19	19	18
10	15	11	7	17
11	20	20	14	2
12	5	18	10	5
13	3	9	3	11
14	13	17	15	3
15	4	6	8	9
16	9	3	20	13
17	2	10	1	6
18	11	7	6	12
19	17	4	18	10
20	18	16	9	4

We used all 9,272 movies to build a latent semantic representation of the types of movies people encounter, but we only used 200 movies as stimuli in the experiment. These movies were chosen in the following way. We selected ten genres (romance, drama, sci-fi, thriller, comedy, horror, animation, fantasy, crime, action), and retrieved the first 20 most popular movies in each genre. This happened in a sequential manner, starting the retrieval with the genres with the lowest number of movies (the final order was horror, romance, animation, fantasy, comedy, thriller, crime, sci-fi, action, drama). The rationale behind this approach was to avoid having relatively unknown movies appearing in a category (in case the well-known ones have already been retrieved as part of another category). We excluded any movies that were part of a sequel if a member of that sequel has already been selected.

Once we knew each movie's loading on the 20 latent dimensions, we could calculate the Euclidean distance between any two of the 200 movies. Based on this calculation, Table 3.2 shows the closest (most similar) and furthest (least similar) five movies for another four selected movies (Interstellar, Inglourious Basterds, The Hobbit: An Unexpected Journey, Star Wars: Episode IV - A New Hope).

This demonstrates that while the algorithm seems to perform well in finding common themes in movies (e.g., the space theme for Interstellar and Star Wars, the war theme for Inglourious Basterds, and the adventure in a magical world theme in The Hobbit), there are a few odd matches. In particular, two movies can have very similar themes, but they might fall into two completely different genres, and thus would never be perceived as similar. A good example is the proximity of WALL-E and Interstellar: while both score high on the outer space theme, WALL-E is an animated family movie, while Interstellar is a sci-fi adventure movie, and therefore they are unlikely to be perceived as similar. Other odd matches include Interstellar – Alien and The Hobbit – Shrek, which further show that latent semantic proximity alone is not enough to create similar movie pairs.

Table 3.2: Closest and furthest five movies for four selected movies based on the Euclidean distance calculated from the LSA solution.

Target	**Interstellar**	**Inglourious Basterds**	**The Hobbit**	**Star Wars**
Closest 1	Alien	Saving Private Ryan	Shrek	Star Trek
Closest 2	Oblivion	Casablanca	The Lord of the Rings	Avatar
Closest 3	WALL·E	The Pianist	Monty Python and the Holy Grail	Guardians of the Galaxy
Closest 4	2001: A Space Odyssey	Full Metal Jacket	Frozen	The Fifth Element
Closest 5	Gravity	Schindler's List	Alice in Wonderland	Oblivion
Furthest 5	Titanic	Man of Steel	Man of Steel	Life of Pi
Furthest 4	Deadpool	Star Trek	Star Trek	Watchmen
Furthest 3	The Wolf of Wall Street	The Dark Knight	The Wolf of Wall Street	Deadpool
Furthest 2	The Dark Knight	The Wolf of Wall Street	The Dark Knight	The Wolf of Wall Street
Furthest 1	Star Wars	Star Wars	Star Wars	The Dark Knight

To mitigate this problem, we decided to also use genre similarity criteria derived from the genre classification information we retrieved from IMDb. This classification contains at least one and at most three main genre types that best describe each movie's type. The simplest approach would be to calculate the number of overlapping genres and rank each movie pair based on this measure. However, there are two problems with this approach. First, the genres used in the IMDb classification are rather crude, since there are only about 18 main genre types. Second, many movies only have one genre assigned to it, decreasing the quality of the genre matching criteria.

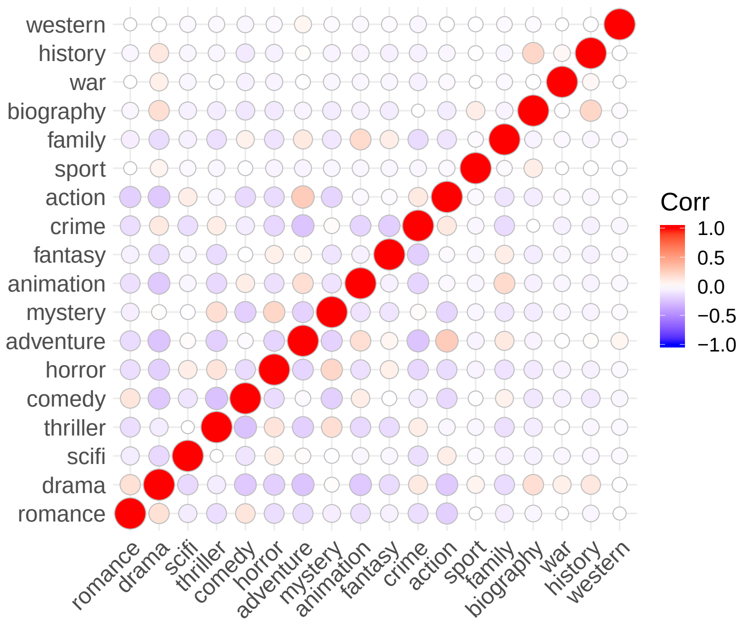
For this reason, we decided to use a different method to calculate the genre similarity between any two movies. Our method consists of three parts, each with a separate score, and the sum of these is the final score, reflecting the genre similarity between each movie pair.

The first part of the score is simply the number of overlapping genres between the two movies. However, as we mentioned above, the genre information is rather scarce for many movies (although every movie has at least one genre), therefore we decided to use the concept of genre similarity as the second criterion. For example, considering a romance, thriller and horror movie, and assuming we do not have any additional genre information on these movies, we would naturally like the thriller and horror movies to be closer to each other than to the romance movie.

To gauge the "proximity" of any two genres, we used genre information from all 9,272 movies in our database to build a genre correlation matrix, which reveals common "genre-mixtures" (see Figure 3.3), that often appear together in genre lists. Then, considering movie pair A and B, we can calculate the number of genres that are positively correlated with the main genre of movie A (the first genre in its genre list) and also appear in the genre list of movie B. For example, using this criteria, the thriller and horror movies would be closer to each other than to the romance movie, since thriller and horror are positively correlated with each other, and both are negatively correlated with romance.

As a real example, consider the movie Fight Club. It only has drama assigned as a category on IMDb. Therefore, based on only the first criterion it is equally close to Lion King, The Lord of the Rings and Se7en (as these all have drama as one of the categories). However, when we add the second criterion, Fight Club has a similarity score of 3 with Se7en (shared genres: drama, crime, mystery), and 1 with Lion King and The Lord of the Rings (the only shared

Figure 3.3: Genre correlation matrix.



category is drama), respectively. Finally, to introduce higher variation in the movie similarity scores, each movie pair was given an additional point if they were retrieved as part of the same genre category from the IMDb website.

At this point we knew the semantic distance and the genre similarity score between each movie pair, and the next step was to construct the target-decoy and target-competitor pairs. We used our LSA- and genre-similarity scores to select pairs of movies that are likely to be perceived as different and similar as target-competitor and target-decoy pairs, respectively. We considered movie pair A and B as different enough to qualify as a target-competitor candidate if 1) the two movies fell in the upper 40% in each other's semantic distance distribution (where a higher distance means the movies are less similar) and 2) their respective genre similarity scores were in the lower 30% of each other's genre similarity score distribution (where a lower genre similarity score means they are less similar).

Considering the variety of the movies people encounter, we would naturally expect to find fewer similar movies than dissimilar ones. Based on this assumption, we used a more stringent set of criteria to find movie pairs similar enough to qualify as target-decoy pairs: the two movies had to fall 1) in the lower 20% of each other's semantic distance distribution and 2) in the upper 15% of each other's genre similarity score distribution. These cut-off values were chosen to provide us with a sufficient number of reasonable matches (601 unique target-decoy and 1,393 target-competitor pairs).

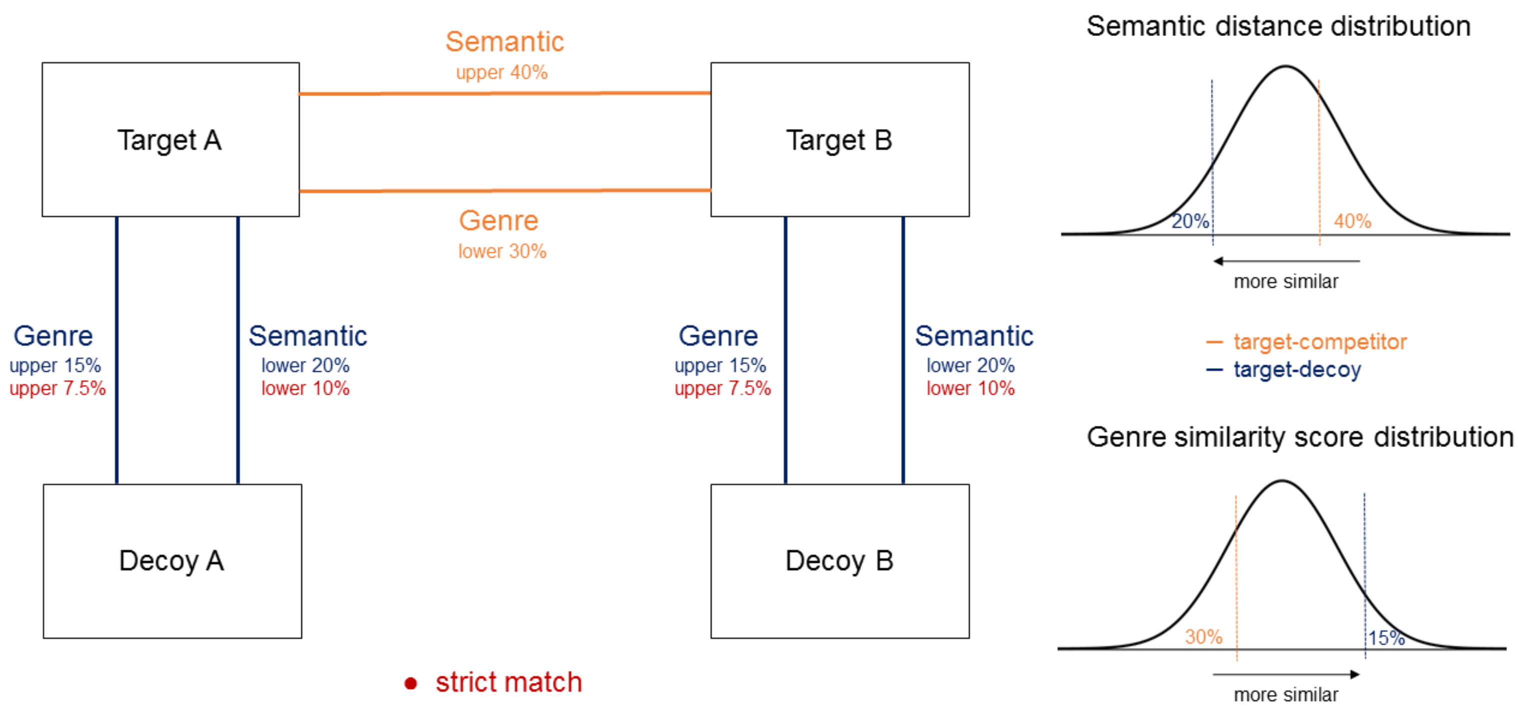
To increase the power of our test, our aim was to create "quadruplets" of movies: two targets A and B, and their decoys A' and B' (as shown on Figure 3.4). A quadruplet can be used to make two triplets: A, B, A' and A, B, B'. That is, each quadruplet consists one target-competitor and two target-decoy pairs. Using our target-decoy and target-competitor pairs, we created 21,886 unique quadruplets. We classified these quadruplets based on the "quality" of their two target-decoy pairs. We called a target-decoy pair a strict match if they fell within 1) the lower 7.5% of each other's semantic distance distribution and 2) within the upper 15% of each other's genre similarity score distribution (these are half the cut-off values we used for creating the target-decoy pairs).

We conjectured that the perceived similarity of the target-decoy pairs will be the most problematic part of recreating an attraction effect (as opposed to finding target-competitor pairs that are perceived as different). This is because we have a highly varied stimuli set with a vast number of storyline-genre combinations, which makes it much harder to find two movies that can be considered similar (especially after eliminating sequels). For this reason, a quadruplet was classified as a high-quality quadruplet if both of the target-decoy pairs in it were classified as strict matches.

3.2.1.3 Experimental procedure

The experiment consisted of three stages: preference rating stage, choice stage, and a similarity rating stage. In the preference rating stage, we asked for par-

Figure 3.4: Quadruplet selection criteria in Experiment 1.



ticipants' subjective evaluations over the 200 movies ("How do you personally rate this movie?") on a scale from 1 (worst) to 7 (best). We also asked whether the participant had seen the movie before. The 200 movies were presented in a random order for each participant. The rating stage took about 15–20 minutes on average. The left panel on Figure 3.5 shows an example of the rating task.

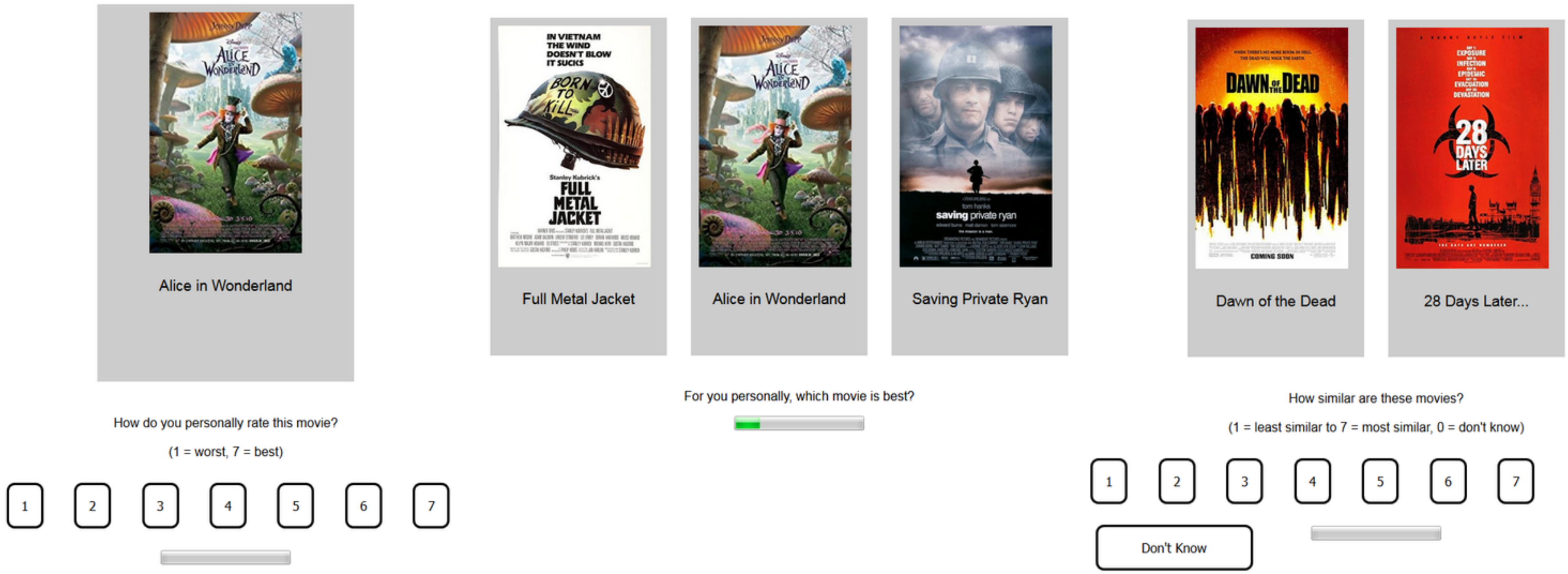
Before the choice stage, we created choice triplets for each participant using the ratings they gave in the preference rating stage in the following way. First, based on the individual ratings from each participant, we identified the subset of quadruplets where: (a) the target and competitor were both rated 4,5,6 or 7, and (b) the two decoy movies were rated at least 3 points lower than the two decoy candidates. Note that we did not require the two decoys in the quadruplet to have the same rating as it would have severely limited the number of quadruplets we could use (e.g., we allowed for quadruplets with ratings 7,7 for the two targets and 4,1 for the two decoys), but we controlled for this difference in our analysis.

We then selected the subset of quadruplets where all of the movies or none of the movies had been seen, to make sure that choice behaviour will not be governed by differences in familiarity with the movies. The result was a bespoke subset of quadruplets for each participant, where the target/competitor movies had the same rating and the decoy movies were rated worse.

However, we did not want the same movie to appear twice as a target/competitor for one participant, and for this reason we used a sequential elimination technique: we first chose the quadruplet with the highest combined target-decoy similarity rating, then eliminated all quadruplets with the same target/competitor movies. We repeated these steps until we had a set of quadruplets with unique target/competitor movies.

We then only invited those participants back for whom we could create at least three high-quality quadruplets this way (corresponding to at least six attraction effect choice triplets where the target-decoy pairs are strict matches).

Figure 3.5: Experiment 1 stages: ratings, choice, similarity ratings.



In the choice stage, participants were presented with the selected movie triplets in a random order and were asked to choose the one they preferred the most out of the three (see the middle panel on Figure 3.5).

Because we had to invite participants back for the choice stage, we collected data in batches of 50 until we had choice data for at least a 100 participants (after all the exclusion criteria had been applied, see section 3.2.1.4).

After the choice stage, those participants whose unique quadruplets included non high-quality quadruplets took part in a third, similarity rating stage. In this stage, they were asked to rate the similarity of the non-strict match movie pairs on a scale from 1 (least similar) to 7 (most similar), where a "don't know" option was also included. The right panel on Figure 3.5 shows an example of the similarity rating task. Information collected in this similarity rating stage was important to ensure the validity of the test.

Overall 322 participants were recruited from the Prolific Academic subject pool whose first language was English and were paid at an hourly rate of £8. We obtained ethics approval from The University of Warwick's Humanities and Social Sciences Research Ethics Committee (reference number: 50/17-18). The description of the experiment asked for participants who were familiar with American movies. We did not collect data on gender, age and ethnicity, as we did not plan to use this information. The typical Prolific Academic user is from the UK, US or Canada ($\sim$ 80%), below the age of 40 ($\sim$ 75%), Caucasian ($\sim$ 79%), female ($\sim$ 60%) and is in full- or part-time employment ($\sim$ 67%). Participants who were invited back after the rating stage were sent an invitation for the choice and similarity rating stage typically three days after they completed the rating stage, and were paid after they completed all three stages.

Out of the 322 participants who completed the rating stage, we could create at least one quadruplet that included movies that had either all been seen or not seen for 262 participants, and for 122 of these participants we could create at least 3 high-quality quadruplets. Out of the 122 participants who were invited back, 114 took part in the choice stage of the experiment.

3.2.1.4 Exclusion criteria

To conduct a rigorous test of the attraction effect, it is crucial that participants take the task seriously and reveal their true preferences. Given that individually rating 200 movies can seem somewhat mundane, we specified a set of exclusion criteria to filter those participants out who did not take the rating task sufficiently seriously. These were the following.

We excluded participants who fell into the fastest 5% of the reaction time distribution, the lowest 5% of the entropy distribution, and the upper and lower 5% of the autocorrelation distribution. Entropy refers to the diversity of the ratings, while autocorrelation takes into account their temporal pattern and measures the extent to which a response depends on previous responses. Thus, this measure aimed to filter out response patterns where participants 1) spent an unusually short time completing the task or 2) did not use the whole of the ratings scale or 3) often gave the same ratings for consecutive movies or 4) were giving ratings randomly.

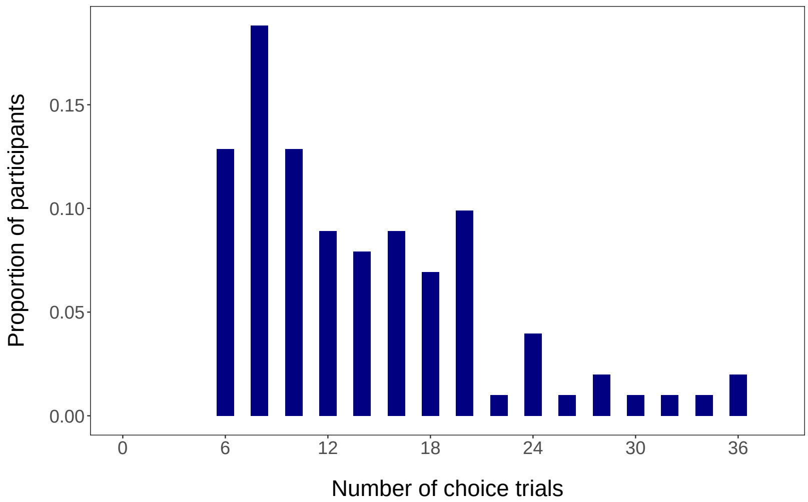
This set of exclusion criteria was validated by a pilot study, where we collected repeated participant ratings for a set of books and found that the subset of participants with a low correlation between repeated ratings ($r < 0.8$) were almost exclusively the ones who were also filtered out by these three criteria. After we applied these exclusion criteria to our sample of 114 participants, we had choice data from 101 participants.

The study design, exclusion criteria and all the analyses were planned and registered before we collected any choice data (for the pre-registration, see Appendix B).

3.2.2 Results

Figure 3.6 shows the distribution of the proportion of participants over the number of choice trials, which, by construction, was always even (since one quadruplet included two choice triplets). The minimum number of choice trials was 6, and the maximum was 36. As it can be seen, 87% of participants were presented with at least 8 choice trials. The average number of choice trials was 14.

Figure 3.6: Distribution of the proportion of participants by number of choice trials in Experiment 1.



As we specified in our pre-registered analysis plan, we excluded any trial in which the participant chose the decoy from the analyses presented below. Figure 3.7 shows that the decoy was rarely chosen, in fact, 61% of the participants never chose it, and less than 4% participants chose it in more than 25% of the trials. This indicates that most participants gave reliable ratings in the first stage, and thus were able to identify the dominated decoy in the choice stage.

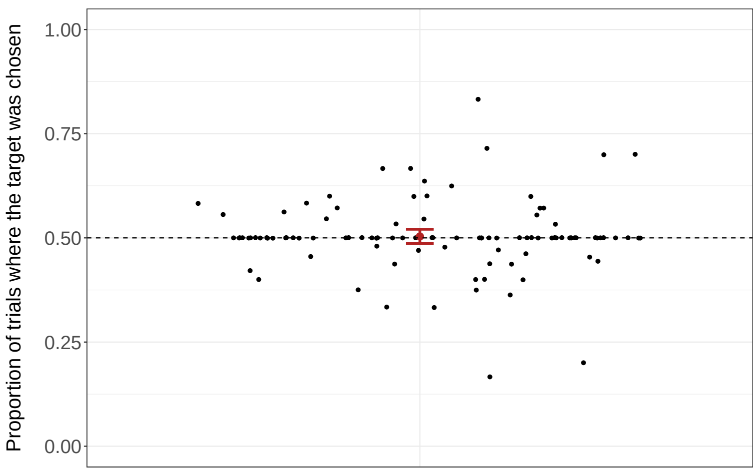
To test for the presence of the attraction effect, we conducted a one-sample t-test to test the hypothesis that the mean of the proportion of trials where the target was chosen is above 0.5. Therefore, evidence for the alternative hypothesis would indicate an increased likelihood of choosing the target item.

For this analysis, we used the triplets created from high-quality quadru-

Figure 3.7: Distribution of the proportion of participants by proportion of trials where the decoy was chosen in Experiment 1.

plets (involving target-decoy pairs that were strict matches). Using a one-sided t-test to test whether participants were more likely to choose the target than the competitor, we found no evidence for the attraction effect (M = 0.50, SD = 0.09), t(100) = 0.39, p = .348. Figure 3.8 shows the distribution of the proportion of trials where the target was chosen, indicating that the vast majority of participants were indifferent between the target and competitor, implying that the presence of the decoy did not affect preferences, M = 0.50, 95% CI [0.49–0.52].

Figure 3.8: Distribution of the proportion of trials where the target was chosen in Experiment 1 (triplets with strict target-decoy pairs only). The red dot and error bars show the bootstrapped mean and 95% CIs.



As specified in our pre-registered plan, we repeated the analysis for triplets with strict target-decoy movie pairs plus triplets created from half the remaining quadruplets with the lowest combined target-decoy semantic distance (we refer to these as the "better half" of the remaining quadruplets), and again found no evidence for the attraction effect (M = 0.51, SD = 0.08), t(100) = 0.28, p = .22 (see Appendix B for the distribution of the proportion of trials where the target was chosen in this subset of data).

Our last planned analysis was a mixed effects logistic regression model with subject-specific random intercepts, where for each trial, we predict the likelihood of choosing the target item with the following explanatory variables: having seen all three of the movies (as opposed to not having seen any of them), target-decoy and target-competitor semantic/genre distance, and the target-decoy rating difference (by the design of the triplet construction method, this was at least 3, and at most 6). Table 3.3 shows the results from this regression.

Table 3.3: Odds-ratios and 95% CIs from a mixed-effects logistic model with subject-specific intercepts, Experiment 1. (T – Target, C – Competitor, D – Decoy)

	Dependent variable:
	Target chosen
Seen all	1.156 (0.844, 1.589)
TC semantic distance	0.980 (0.878, 1.093)
TD semantic distance	1.015 (0.910, 1.131)
TC genre distance	0.993 (0.890, 1.107)
TD genre distance	0.972 (0.872, 1.084)
TD rating difference	0.908 (0.789, 1.046)
Intercept	0.992 (0.628, 1.566)
Observations	1,418
Log Likelihood	−978.245
Akaike Inf. Crit.	1,972.490
Bayesian Inf. Crit.	2,014.546
Note:	*p<0.1; **p<0.05; ***p<0.01

As it can be seen, familiarity with the movies, preference difference between the target and decoy movies, and the semantic/genre distance between the target-decoy and target-competitor pairs do not affect the likelihood that the target will be chosen. Based on this model, we can predict the probability of choosing the target in an "ideal" attraction effect situation, where the target-decoy rating difference is at its maximum (6), all the movies had been seen in the triplet, and the genre/semantic difference between the target-decoy and target-competitor pairs are at their respective minimum and maximum. This predicted proportion[1] is only 43%, 95% CI [30%–57%], which further shows the lack of an increased tendency to choose the target even in choice situations where the attraction effect is most likely to be observed.

We can only speculate about the underlying reason(s) why we did not see an increased preference for the target in our experiment. One possible explanation for our results is that the attraction effect is simply absent in settings where choice options are complex objects, as argued by Frederick et al. (2014) and Yang and Lynn (2014). While this is one possible explanation for our results, we cannot be certain unless we can safely claim that our experimental methodology managed to satisfy all the criteria required for a stringent test of the attraction effect. Testing the attraction effect with naturalistic stimuli is a complex task, as it poses several challenges to the standard experimental methodology commonly used in research focusing on this decision bias.

First, naturalistic objects differ from numerical stimuli as we cannot infer the relative value of each option "on the spot". To circumvent this issue, we added a rating stage before the actual choice task, and there was at least a day's difference between the completion of the two tasks. One natural concern could be that this might have reduced the reliability of ratings. However, the results have shown that participants were (1) indifferent between the options they have previously rated the same and (2) rarely ended up choosing the dominated option. These results indicate that participants gave honest answers in the rating stage and their preferences did not change over time.

[1]Based on a logistic regression without mixed effects.

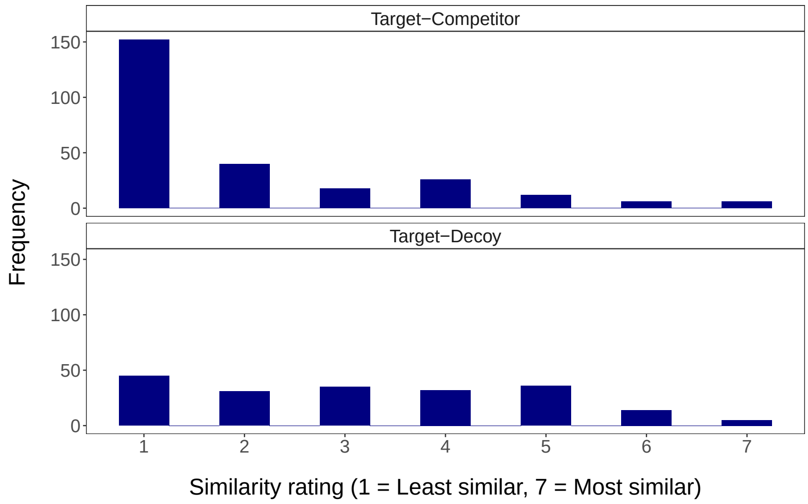
Second, one must make sure that participants are sufficiently interested in the stimuli. With numerical stimuli, interest can be sustained by using an incentive-compatible experimental design. To ensure these criteria are met in our experiment, we used the most popular movies on IMDb, employed a very strict exclusion criteria to exclude participants who did not take the task seriously, and also controlled for familiarity with the choice options, but this did not modulate the strength of the effect.

Third, the similarity comparison process might differ between options with numerical and naturalistic attributes. The perceived similarity of the target and decoy is a crucial assumption, as many choice models rely on it to be able to produce the attraction effect (see subsection 3.1.1). To ensure maximal similarity between the target and decoy, we used all available information on our movies, including storyline, actors, and genres, to create the target-decoy and target-competitor pairs, and conducted separate analyses based on the quality of the match. However, this still does not ensure that participants perceived the target-decoy pairs as similar and the target-competitor pair as different.

Fortunately, we collected information on perceived similarity for a small subset of movie pairs that did not qualify as strict matches based on our selection criteria, which allows us to estimate the extent of the problem. Specifically, we have at least one similarity rating for 184 unique target-decoy and 203 unique target-competitor pairs, which constitute 16% and 18% of all unique choice triplets, respectively. Figure 3.9 shows the distribution of the similarity ratings for these target-decoy and target-competitor pairs.

It seems that while most target-competitor movie pairs were given low similarity ratings, showing that they were perceived as markedly different (as we hoped), participants generally did not perceive our target-decoy pairs to be very similar. Although it is worth keeping in mind that we asked for similarity ratings for this subset of movie pairs exactly because we were worried that they might not be perceived as different/similar enough (therefore, it is likely that

Figure 3.9: Distribution target-competitor and target-decoy similarity ratings in Experiment 1.



target-decoy pairs that were strict matches were perceived as more similar than the non-strict matches), but the very low target-decoy similarity ratings still cast doubt on the general efficacy of our triplet selection criteria.

While we predicted in advance that creating pairs of similar movies will be the most challenging part of the experimental design, the lack of perceived similarity is highly problematic when designing a test of the attraction effect. Specifically, if the proximity of the target and decoy are not recognised, then the choice task simplifies to a situation with two, equally good, distinct options, and a third, significantly inferior option. In this case, we would expect that the decoy will not have any effect of the subjective valuation of the two, equally desirable options and that they will be chosen equally frequently, which is exactly the pattern we have seen in our results.

3.2.3 Discussion

Our aim in Experiment 1 was to test the attraction effect using naturalistic, real-world stimuli. Our results suggested that adding a dominated decoy to a binary choice set did not increase preference for the dominating target. We are 95% confident that the true proportion of trials where the target was chosen is between

49% and 52%. In addition, the tendency to choose the target was not affected by participants' familiarity with the movies, the target-decoy rating difference, or the semantic/genre proximity of the target-decoy and target-competitor pairs.

Furthermore, the distribution of the target-decoy similarity ratings from a subset of triplets show that participants rarely perceived the target-decoy pairs as similar. This indicates that (at least) some of the triplets we created did not manage to invoke an attraction effect choice situation.

Why did our stimuli selection method fail to produce movie pairs that are similar? As we mentioned in section 3.2.1.2, while the latent semantic solution was successful in retrieving common themes and aspects of each movie's storyline, genre differences are not captured well in this framework (as a certain theme can appear in very different contexts), and it is possible that they are central to the similarity judgement process.

While our genre matching criteria were supposed to account for this problem, it is possible that matching movies along 18 genre types simply does not capture subtle genre differences that are crucial in judging the similarity between two movies. This problem is further exacerbated by the fact that many movies have only one category assigned on IMDb, making it much harder to establish movie similarity along the genre dimension.

For example, The Life of Brian has comedy as the only category assigned, whereas Monty Python and The Holy Grail has adventure, comedy and fantasy as genre categories, which means that our algorithm does not identify these two movies as similar enough along the genre dimension, despite the fact that they are both Monty Python movies (and would probably be considered as very similar by most people).

To summarise, in Experiment 1, we find no evidence for the attraction effect. However, the generally low target-decoy similarity ratings suggest that in a significant proportion of the trials, we simply failed to create an attraction effect type situation. To conduct a rigorous test of the attraction effect using the same stimuli, this problem needs to be accounted for.

3.3 Experiment 2

Experiment 2 also aimed to test the attraction effect with the same naturalistic choice options, while addressing the problems encountered in Experiment 1. Specifically, to ensure that target-decoy pairs were perceived as similar and that target-competitor pairs were perceived as different, we used similarity ratings from a separate, independent group of participants to help construct choice triplets. We also collected similarity ratings on every target-decoy and target-competitor pair encountered in the choice stage immediately after the choices, which enables us to directly test how the strength of any attraction effect depends on the perceived similarity of the target-decoy and target-competitor pairs.

3.3.1 Method

We followed a slightly simplified version of the method described in section 3.2.1. Similarly to Experiment 1, the study design, exclusion criteria, and all the analyses were planned and registered before we collected any choice data (for the pre-registration, see Appendix B).

3.3.1.1 Stimuli

Using the same movie selection procedure (retrieving the most popular movies from the 10 selected genres) we doubled the size of the stimuli set, amounting to 400 movies. We extended the stimuli set in the hope that this will increase the likelihood of finding pairs of similar movies (while still using movies our participants are likely to be familiar with).

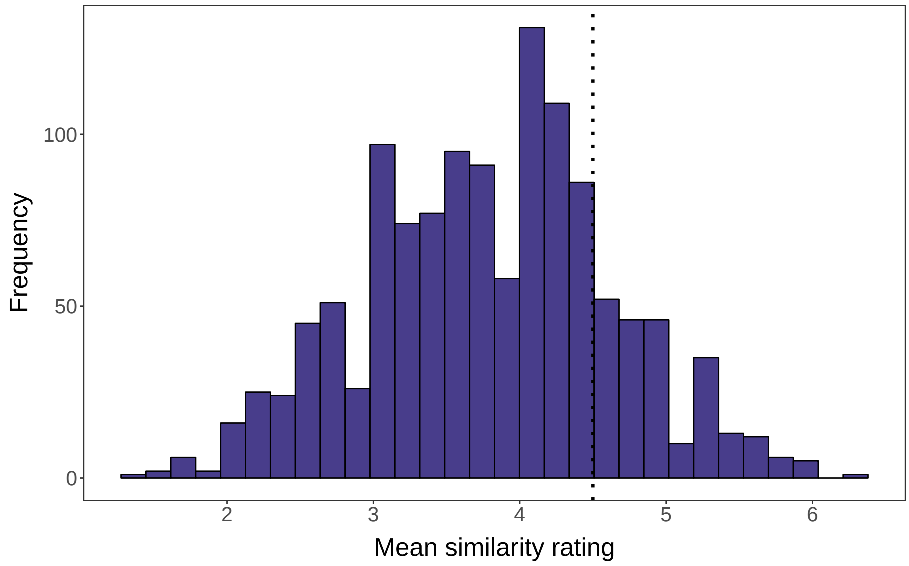
3.3.1.2 Choice set selection

We mostly used a genre matching criteria to identify potential target-decoy and target-competitor pairs. To do this, we used additional genre information from the allmovie.com website. The genre information on allmovie.com is much richer than on IMDb: compared to the 18 genre categories on IMDb, there are 156 genre and sub-genre categories, capturing many important aspects of the movies. For example, using this criteria, The Life of Brian (main genre: comedy, sub-genres: parody/spoof, absurd comedy, religious comedy) and Monty Python and The Holy Grail (main genre: comedy, sub-genres: absurd comedy, parody/spoof, slapstick) are very close to each other, just as we would expect. Using this rich genre information, we created a movie by movie (400 x 400) matrix, where each cell was the number of overlapping genre categories between the two movies.

We then paired up movies that were both scoring highest in each other's genre score distribution, creating 2,271 target-decoy candidates. We also added 806 pairs obtained from the mutually closest 10% of movies based on the latent semantic solution with 20 and 74 latent dimensions (the 806 pairs were the ones that appeared in both solutions). The rationale behind this approach was to capture movie pairs that are very close to each other in terms of the story themes, but are not the closest on the genre dimension. Overall, we had 3,011 unique movie pairs at this point.

Our next task was to reduce the size of this list by selecting the most similar movie pairs. This was done manually by two researchers, who gave a similarity rating between 1–7 for each movie pair (1 − least similar, 7 − most similar). We only kept the movie pairs that had a similarity rating above 1, resulting in 1,242 target-decoy candidates. We then divided the 1,242 pairs into six groups of 207 pairs, and ran an independent pilot study where we asked 60 participants to rate the similarity of a randomly chosen group of movie pairs, obtaining 10 independent similarity ratings for each of the 1,242 target-decoy candidates. Participants rated the similarity of each movie pair on a 1–7 scale, where 1 is the least and 7 is the most similar, and a “Don't know” option was also available. Figure 3.10 shows the distribution of the average similarity ratings for each movie pair. We retained information about the mean similarity of each target-decoy pair, which we later used in the choice set selection stage,

Figure 3.10: Distribution of the average similarity rating for each target-decoy candidate.



to discriminate between quadruplets that contain the same target or competitor movies.

We then decided to only use movie pairs with ratings that are equal to or higher than 4.5, which corresponds to the upper 20% of the similarity rating distribution (253 movie pairs). We hoped this procedure would ensure that our target-decoy candidate pairs are perceived as similar by most participants.

Once we had a sufficient number of movie pairs that we believed were reasonably similar, the next step in creating the quadruplets was to create target-competitor pairs. To do this, we first created a target-decoy pair by target-decoy pair matrix (253x253), where each cell was the number of overlapping genre categories between the two movie pairs. For example, considering the comparison between target-decoy candidate 1 (consisting of movie A and movie B) and target-decoy candidate 2 (consisting of movie C and movie D), we summed the number of genre overlaps between movies A–C, A–D, B–C and B–D.

Then we created the quadruplets by combining each target-decoy pair with all the target-decoy pairs that it had no common genre categories with, and taking the unique combinations of the four movies. This resulted in 20,022 quadruplets, created from 231 unique movies.

3.3.1.3 Experimental procedure

The experimental procedure was almost identical to the one described in Experiment 1 (see section 3.2.1.3). We first asked participants to provide preference ratings for 231 movies and to specify whether they had seen the movies. After we collected ratings data for the next 50 participants, we identified the quadruplets in which 1) all the movies had either been seen or not seen, 2) the two targets had the same rating and 3) the rating difference between the two targets and their respective decoys was at least 3. Using the same sequential selection procedure (using the sum of the two similarity ratings as an indicator for the quality of the choice set), we then obtained a subset of these quadruplets with unique target and competitor movies. We invited back those participants for whom we could create at least three quadruplets.

After the choice stage, we asked each participant to rate the similarity of each target-decoy and target-competitor pair they encountered in the previous stage. While asking participants to rate the similarity of all target-decoy and target-competitor pairs is costly and time-consuming, this information enabled us to account for individual heterogeneity in similarity perception and make sure that our choice triplets represented an attraction effect choice situation for each participant.

We recruited 297 participants using Prolific Academic. Out of the 297 participants that completed part 1, we could create at least one quadruplet that included movies that had either been all seen or not seen for 249 participants, and for 179 of these participants we could create at least 3 quadruplets. Out of the 179 participants who were invited back, 152 took part in the second stage of the experiment.

3.3.1.4 Exclusion criteria

We used the same exclusion criteria described in section 3.2.1.4. After we applied these exclusion criteria to our sample of 152 participants, we had choice data from 135 participants.

3.3.2 Results

Figure 3.11 shows the distribution of the proportion of participants by the number of choice trials. Similarly to Experiment 1, all participants completed an even number of choice trials, the lowest number of choice trials was 6, and the highest was 54. The average number of trials was 16, and 84% of participants were presented with at least 8 choice trials.

Figure 3.11: Distribution of the proportion of participants by number of choice trials in Experiment 2.

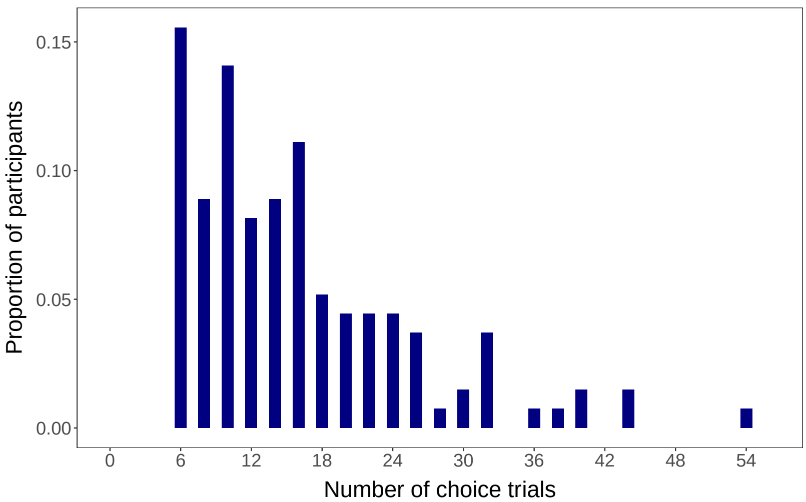
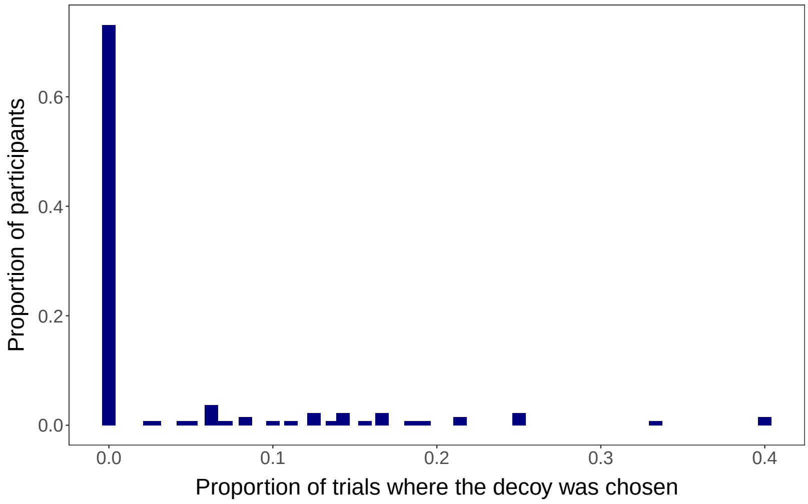


Figure 3.12 shows that similarly to Experiment 1, the decoy was again rarely chosen. In fact, 72% of the participants have never chosen it, and only 2% of participants have chosen it in more than 25% of the trials. This indicates that participants were again able to identify the dominated decoy in the choice stage.

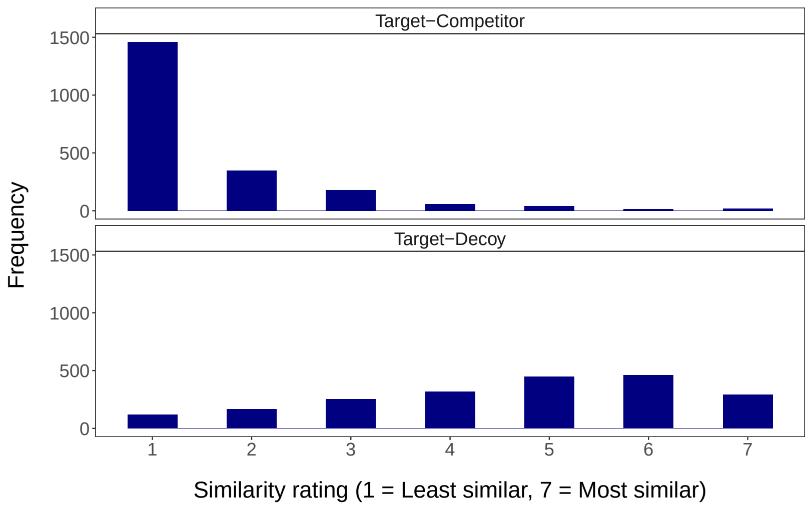
Before analysing choice behaviour, it is instructive to plot the target-

Figure 3.12: Distribution of the proportion of participants by proportion of trials where the decoy was chosen in Experiment 2.



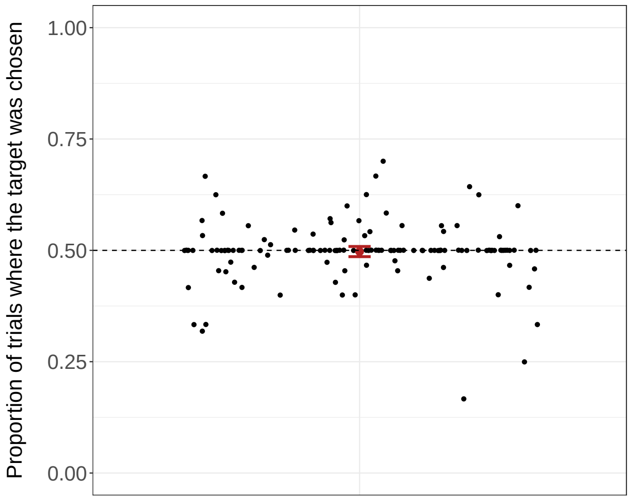
decoy and target-competitor similarity rating distributions, to see if we managed to create more similar target-decoy pairs in Experiment 2. Figure 3.13 shows these distributions, and we can see that the overwhelming majority of target-competitor pairs were perceived as not similar, while the majority of target-decoy pairs were perceived as similar, which shows that we indeed managed to improve the perceived similarity of the target-decoy pairs.

Figure 3.13: Distribution target-competitor and target-decoy similarity ratings from Experiment 2.



To test for the presence of the attraction effect, we again conducted a

Figure 3.14: Proportion of trials where the decoy was chosen in Experiment 2. Each dot is a participant and the red dot and error bars show the bootstrapped mean and 95% CIs.



one-sample t-test to test the hypothesis that the mean of the proportion of trials where the target was chosen is above 0.5 (indicating an increased likelihood of choosing the target item). The results are very similar to those from Experiment 1: we again found no evidence for the attraction effect (M = 0.49, SD = 0.07), t(134) = -0.44, p = .669. Figure 3.14 shows the distribution of the proportions of trials where the target was chosen, which again indicates that participants were indifferent between the target and the competitor, M = 0.50, 95% CI [0.49–0.51].

As specified in our pre-registered analysis plan, we also ran a mixed effects logistic regression with subject-specific intercept to investigate how the target-decoy and target-competitor similarity ratings and familiarity of with the movies affect the likelihood of choosing the target. Table 3.4 shows the results from this regression, where contrary to our expectations, none of the explanatory variables seem to modulate the strength of the attraction effect.

We can again estimate the probability of choosing the target in an "ideal" attraction effect scenario, where the target-decoy pair is perceived as the most similar, whereas the target-competitor pair is perceived as the least similar. This predicted proportion[2] is again only 41%, 95% CI [31%, 52%], showing that

[2]Based on a logistic regression without mixed effects.

Table 3.4: Odds-ratios and 95% CIs from a mixed-effects logistic model with subject-specific intercepts, Experiment 2. (T – Target, C – Competitor, D – Decoy)

	Dependent variable:
	Target chosen
Seen all	1.122 (0.731, 1.729)
TC similarity rating	0.967 (0.872, 1.073)
TD similarity rating	0.926 (0.835, 1.027)
TD rating difference	0.935 (0.837, 1.044)
Intercept	1.062 (0.598, 1.882)
Observations	1,541
Log Likelihood	−1,064.822
Akaike Inf. Crit.	2,141.644
Bayesian Inf. Crit.	2,173.686
Note:	*p<0.1; **p<0.05; ***p<0.01

even when the target-decoy and target-competitor pairs are perceived as very similar/different, we do not see an increased tendency to choose the target (if anything, we see a slight tendency to choose the competitor).

3.3.3 Discussion

Experiment 2 aimed to test the attraction effect with complex, naturalistic stimuli, while accounting for the methodological shortcomings of Experiment 1. Based on the results from Experiment 1, we hypothesized that we did not find evidence for the attraction effect in Experiment 1 because participants generally did not perceive the target-decoy pairs as similar.

For this reason, in Experiment 2, we used a rich set of genre information to create similar movie pairs. In addition, we collected similarity ratings for each target-decoy and target-competitor pair participants encountered in the choice stage. This allowed us to see how the strength of the attraction effect depends on the perceived similarity of the movie pairs.

The results are broadly similar to those obtained from Experiment 1. That is, we have found that the presence of a dominated decoy does not alter preferences between the target and the competitor (we are 95% confident that the true proportion of trials where the target was chosen is between 49% and 51%). In addition, to our surprise, the tendency to choose the target was not affected by the perceived similarity of the target-decoy and target-competitor pair.

3.4 General Discussion

In Experiment 1 and 2, we explored the strength of the attraction effect in a naturalistic choice context. While previous research had shown that the effect is rather robust in choice tasks where the attributes have a numerical representation, its relevance has been questioned in choice contexts involving naturalistic options.

To conduct a test of the attraction effect using complex stimuli, we used movie posters as real-world choice options to create target-competitor-decoy choice triplets. In Experiment 1, we did not find any evidence for the attraction effect, but we also found that participants did not perceive our target-decoy pairs as similar, which cast doubt on the validity of our test. To address this problem, we conducted the same test with an improved methodology in Experiment 2. We believe that Experiment 2 is the first rigorous investigation of this research question: we designed it carefully to address all the criticisms raised in connection with the study by Frederick et al., where they used similar stimuli.

First, our experimental design ensured participants' indifference between the target and the decoy, maximising the probability that choices will be constructed on the spot (rather than through relying on strong prior preferences), and an attraction effect will occur. While one could argue that mnemonic processes arising from familiarity with the stimuli can alter preferences in the choice stage, we still did not detect an increased tendency to choose the target when participants were not familiar with the movies.

Second, we have strong evidence that the dominance relationship was perceived in our experiment. The target-decoy similarity ratings confirmed that our careful target-decoy selection process indeed managed to produce movie pairs that were perceived as similar. In addition, we ensured that the decoy was always rated at least 3 units lower than the target (and the competitor). Accordingly, the decoy was only chosen in 4.3% of the trials, which clearly shows that participants were able to spot and avoid the dominated alternative.

Third, by creating bespoke choice triplets based on preference ratings, we avoided individual heterogeneity in preferences to act as a potential confound. In addition, we ensured that the decoy was not too desirable in comparison to the target. We also used a set of strict exclusion criteria to filter out participants who did not take the task sufficiently seriously, and with an average of 16 choice trials per participant, we avoided participant fatigue.

Finally, in our analysis we controlled for familiarity with the choice options, perceived similarity of the target-decoy and target-competitor pair, and relative preference between the target and the decoy, but we found that none of these modulated the strength of the attraction effect.

The results from this rigorous test duplicated the findings from Experiment 1: the presence of the decoy in the choice set did not alter preferences between the target and the competitor. Therefore, we can conclude that even when we accounted for the criticisms in connection with Frederick et al. and Yang and Lynn's studies, we still did not find the attraction effect when options had complex, naturalistic attributes. This illuminates the possibility that the discrepancy stems directly from the stimuli presentation, which might give rise to fundamentally different comparison processes.

Unfortunately, Frederick et al. did not elaborate on the potential reasons for their results. They briefly conclude that it is the numerical nature of the attribute dimensions that gives rise to the attraction effect, but we think this is a simplification. In fact, it has been shown on numerous occasions that the attraction effect is also present in settings where options do not have numerical attributes.

For example, Choplin and Hummel (2005) have demonstrated the effect in a similarity judgment task with unidimensional stimuli (by varying the aspect ratio of circles and the length of lines). Maylor and Roberts (2007) have found the attraction effect in a memory judgment task. Farmer, El-Deredy, Howes, and Warren (2015) have demonstrated the effect in motor planning decisions. In addition, the attraction effect has also been found in experiments using perceptual stimuli. In Trueblood et al., participants were presented with rectangles of differing height and width, and had to select the rectangle with the largest area. Other studies have used the same rectangle stimuli and found the attraction effect in children (Zhen & Yu, 2016) and monkeys (Parrish et al., 2015).

Interestingly, Frederick et al. presented the only experiment we know about that did not find the attraction effect using Trueblood et al., 2013's perceptual stimuli. However, important differences exist between the two experimental methodologies, demonstrated by the fact that in Frederick et al.'s study, the decoy's choice share was double of that reported in Trueblood et al.'s experiment. Nevertheless, we think replications are key in understanding the boundary conditions of the attraction effect. We are aware of the fact that there is an unfortunate tendency in psychology not to report failed replication attempts (Ingre & Nilsonne, 2018), which hinders our understanding of how the strength of the attraction effect depends on the stimuli characteristics.

We find these contradictory results interesting, as they raise the possibility that it is not the numerical nature of the attribute dimensions, but some other attribute characteristic that underlies the attraction effect. We will investigate this question in Chapter 4.

Chapter 4

Exploring the factors underlying the attraction effect

4.1 Introduction

In Chapter 3, we conducted two experiments to probe the robustness of the attraction effect with complex, naturalistic stimuli. This research question arose from the debate on the real-world relevance of the attraction effect, which started with the study of Frederick et al. (2014) that described numerous failed attempts to replicate the attraction effect in choice contexts where the options were naturalistic objects. They drew the conclusion that the attraction effect is only present in choice settings where the options have numerical attributes. The findings from our experiments confirmed Frederick et al.'s results in that we have not found any evidence for the attraction effect using movie posters as naturalistic stimuli.

However, as discussed in Chapter 3, there is substantial evidence that the attraction effect is *not* restricted to choice settings where the options have numerical attributes. In value-based decision making, Farmer, Warren, El-Deredy, and Howes (2017) have demonstrated the attraction effect using perceptual representation of gambles. In one version, they adopted the rectangle stimuli of Trueblood et al. (2013) to represent the probability and the nominal amount to be won (therefore, the area of each rectangle corresponded to the expected value of the gamble). In another experiment, the probability of the gamble was represented by the proportion of randomly distributed coloured squares in a 10x10 grid, whereas the nominal amount was represented as the proportion of coloured squares in another 10x10 grid.

While these perceptual tasks are clear departures from the commonly used numerical attributes format, they are still very similar to it in one respect. Specifically, in both versions, the two attribute dimensions are spatially separated, and thus can be attended independently. We argue that independent processing of attribute dimensions can be key in explaining why the attraction effect is so elusive in some choice contexts. To design an experiment that tests this hypothesis, we draw on insights from decades of research on the perception and processing of multiattribute stimuli.

4.1.1 Integral and separate dimensions

The idea that multiattribute stimuli can be classified based on the perceived separability of the attribute dimensions first emerged from research focusing on similarity judgments using the method of multidimensional scaling (MDS; Kruskal, 1964; Shepard, 1980). In essence, MDS can be used to build a low-dimensional geometrical representation of the perceived similarity between pairs of objects (Hout, Papesh, & Goldinger, 2013). In this representation, the objects are points in a Cartesian coordinate system, and the distance between them corresponds to their perceived similarity (so that more similar objects are closer to each other). This distance can be calculated in various ways, of which the Euclidean distance metric, $\delta = \sqrt{\delta_x^2 + \delta_y^2}$ is the most intuitive and well-known, and was consequently used in the first applications of MDS to similarity judgments.

However, Attneave (1950) challenged the appropriacy of the Euclidean distance metric. He proposed that an alternative distance measure, the city-block metric, $\delta = \delta_x + \delta_y$ provides a much better fit for the similarity judgment data he collected in his experiments, where he asked participants to make similarity judgments between parallelograms, squares and triangles with varying dimensions, including size, tilt, and colour. According to the city-block metric, the perceived overall distance is simply the sum of the distances along the attribute dimensions.

In contrast with Attneave's results, Torgerson (1958) found strong support for the Euclidean metric in an experiment where participants were asked to provide similarity judgments for Munsell colour chips that differed in brightness and saturation. These contradictory results led researchers to examine how the perceived similarity of two objects depend on their attribute characteristics.

In fact, the Euclidean and city-block distance metrics reflect two fundamentally different ways of perceiving object similarity. Specifically, the Euclidean distance metric is invariant to axis rotation, while the city-block metric is not. This means that objects whose pairwise similarity can be best described by the city-block metric have “privileged” psychological dimensions, whereas the Euclidean metric is more appropriate for objects that are processed “holistically”, where the underlying attribute dimensions may not even be perceived separately (Shepp & Ballesteros, 1989).

Shepard (1964) argued that these differences can be explored by analysing how participants justify their similarity judgments for these two classes of stimuli. For example, when the attributes can be perceived independently (stimuli that he called *analyzable*), participants have almost always referred to the distinct attribute dimensions when describing differences. However, when the stimuli are processed holistically (for dimensions he called *unitary*), participants tended to describe the difference along one, “combined” dimension, indicating that the two attribute dimensions cannot be perceived separately (and that the participant might not even be aware that there are multiple underlying dimensions).

Lockhead (1966) investigated how the effect of dimensional redundancy differs for stimuli with *separable* and *integral* dimensions (these are the respective equivalents of *analyzable* and *unitary* in Shepard’s terminology). A redundancy gain is said to occur when performance in a selective attention task (most typically discrimination, detection, or categorization) is improved (faster and/or more accurate responses) when the two attribute dimensions vary in a correlated manner. Previous research found redundancy gains for integral dimensions (Munsell colour chips; Eriksen & Hake, 1955), but not for separable dimensions (visual positions of X’s and O’s; Garner & Lee, 1962). Lockhead argued that the redundancy gain occurs because integral attributes cannot be attended separately, and that the concept of redundancy gain should be key in the definition of integral attribute dimensions.

Drawing on Lockhead’s findings, Garner and Felfoldy (1970) demonstrated that integral dimensions give rise to redundancy gains and interference effects within the same experimental task. Interference effects (slower and less accurate responses) can be observed if the two integral dimensions are orthogonal (vary in an uncorrelated manner). They proposed a new definition for integral attribute dimensions. According to this, dimensional integrality (1) can be best described by the Euclidean distance metric, (2) produces redundancy gains if the dimensions vary in a correlated fashion, and (3) results in interference effects when the two dimensions are orthogonal. The experimental paradigm where these assumptions can be tested with a control, correlated, and filtering condition has later became known as the Garner interference task (Burns, 2014).

More recent research focused on how the processing of stimuli with integral and separable dimensions differ in perceptual categorization tasks. These studies have typically analysed reaction times and choice probabilities in sequential sampling modelling frameworks. Little, Nosofsky, and Denton (2011) investigated whether the processing of multiattribute stimuli depends on the spatial separability of the attribute dimensions. They found that when the attribute dimensions are spatially separated (e.g., the base width of a lamp and the curvature of its top piece), they are processed in a sequential, serial manner. However, when the attribute dimensions are spatially overlapping, such that both attributes can be attended simultaneously (e.g., the colour of a rectangle and the position of an inset bar), a mixture of serial and parallel processing of dimensions occur. In a follow-up study, they found strong evidence that stimuli with integral attributes (e.g., brightness and saturation) are processed in a coactive fashion, where the attribute information is combined into a single processing channel (Little, Nosofsky, Donkin, & Denton, 2013), as opposed to the parallel and serial model, where there are two independent accumulation processes for the two attribute dimensions.

While caution is warranted when applying insights from perceptual categorization to value-based decision making due to the inherent differences between the two choice tasks, if object processing shares commonalities in the two contexts, it bears significance for any research that aims to investigate the cognitive mechanism underlying the attraction effect in preferential choice. Crucially, most sequential sampling models that are able to accommodate the attraction effect rely on the assumption of attribute-wise processing of choice options. In addition, eye-tracking evidence suggests that single-attribute pairwise comparisons play a key role in the attraction effect (Noguchi & Stewart, 2014).

To elucidate whether it is the separable nature of the choice attributes that gives rise to the attraction effect in value-based decision making, we conducted an experiment to test whether the strength of the effect depends on stimulus presentation. Participants first had to learn a valuation rule in a learning stage, and they were instructed to use this rule in the choice stage where they were presented with attraction effect choice triplets created form two kinds of artificial stimuli (separable and integral version).

4.2 Experiment 1

4.2.1 Method

In this experiment, our aim was to investigate the hypothesis that the strength of the attraction effect depends on the separability of the attribute dimensions. To this end, we created two versions of the same stimuli: a “traditional” version, where the two attribute dimensions have numerical values (numerical condition), and a perceptual version, where the attribute dimensions are integral, and thus cannot be attended independently (pictorial condition). We expected to see the attraction effect in the numerical condition, where the options can be compared along an attribute dimension, but not in the pictorial condition, where the options can only be processed holistically.

Designing a value-based choice experiment where the attribute dimensions are integral and preferences are monotone and continuous is challenging.

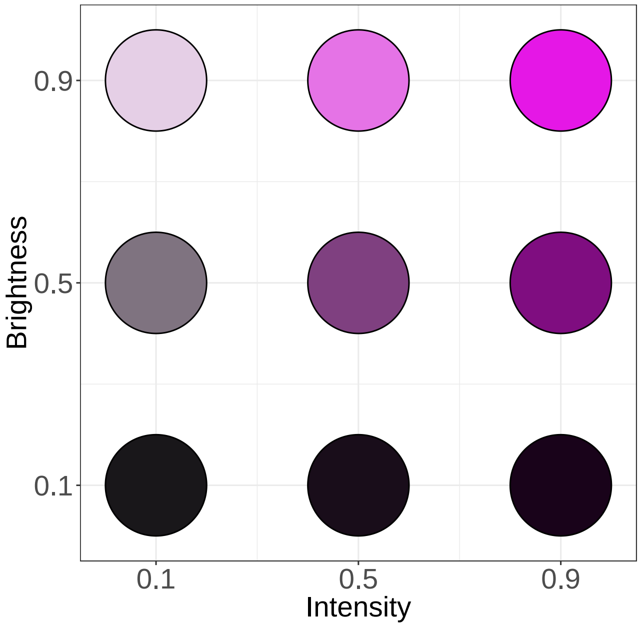
To be consistent with previous research, we chose the most well-known integral stimuli in the literature, Munsell colours with a fixed hue, but varying brightness and intensity. Numerous studies have established that these attribute dimensions are integral, and therefore are processed holistically (e.g., Eriksen & Hake, 1955; Garner & Felfoldy, 1970).

We are only aware of a few studies that used perceptual stimuli in a preferential choice context. In these, the value of the options naturally depended on the perceptual representation of the attribute dimensions (e.g., height and width of rectangle, where the value is given by the area; Trueblood et al., 2013). However, when the attributes are integral, it is much harder to "induce" preferences based on a valuation rule, due to the very nature of the stimuli (i.e., the inaccessibility of the independent dimensions). To overcome this difficulty, we first established a valuation rule that assigns a nominal value to each option, based on the two attribute dimensions. Participants had to learn this valuation rule in a learning stage that preceded the choice stage. They were then instructed to use the valuation rule when making decisions in the choice stage. While inducing preferences through the learning stage is somewhat artificial, this method ensured that participants had continuous, monotone and homogeneous preferences over our stimuli set.

4.2.2 Stimuli

In colour theory, there are various ways to describe colours, of which the red-green-blue (RGB) colour model is perhaps the most well-known (Shipman, 2012). We created our stimuli using the hue-saturation-value (HSV) colour model. We will refer to saturation as intensity, and to value as brightness. In the HSV model, hue corresponds to the pure colour component of a colour, and is measured as the angle around the colour wheel, ranging from 0° to 360°, with cyan at 180°, and red at both ends of the wheel. For this entire experiment, we fixed the value of hue at 300°, resulting in a purple colour at the mean values of intensity and

Figure 4.1: Illustration of the intensity-brightness dimensions with hue fixed at 300°



brightness. Intensity and brightness both vary between 0 and 1. As shown by Figure 4.1, if brightness is set to 0.5, and hue is fixed at 300°, an intensity level of 0.1 corresponds to a completely grey colour, whereas a value of 0.9 results in an intense purple colour. Equally, if intensity is set at 0.5, and hue is fixed at 300°, a brightness level of 0.1 corresponds to a completely black colour, whereas a value of 0.9 results in a light purple colour.

Our stimuli were teapots with varying brightness and intensity levels. Participants were told that the levels of these two attributes determined each teapot's value. In the numerical condition, the brightness and intensity values were displayed numerically, whereas in the pictorial condition, it was the colour of the teapot that determined the value. Importantly, the value of the choice options were calculated using the same rule in both conditions.

4.2.2.1 Choice triplet selection process

Previous research has shown that one unit change in brightness is perceptually equivalent to a two unit change in intensity (Newhall, 1940), and we took this into account when constructing the rule that determines the value of each option based on the corresponding brightness and intensity values. This relationship is captured by the red line on Figure 4.2, which is defined by the equation

$$Brightness = 0.5 \cdot Intensity + 0.25. \tag{4.1}$$

All stimuli were required to fall within the boundaries of the black polygon displayed on Figure 4.2. The shape of the polygon ensured that intensity and brightness were restricted to fall between 0.05–0.95 and 0.15–0.85, respectively. This was necessary to avoid extreme attribute values that could hinder the perception of the other attribute dimension (e.g., a very low brightness would translate into a black colour, regardless of the intensity value).

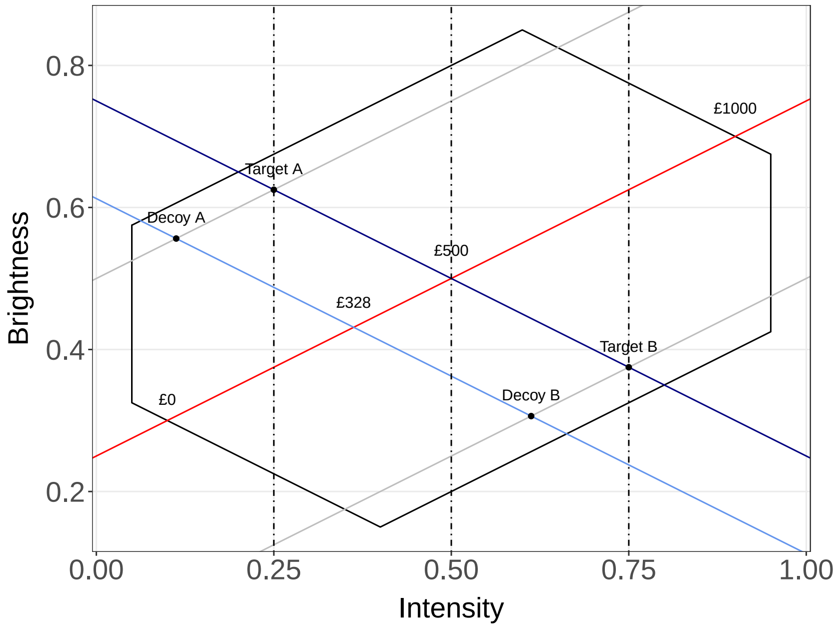
To create an attraction effect choice triplet, we first determined the position of the target and competitor options within the polygon, followed by the decoy. Then, we calculated the nominal value of each of the three options using Equation 4.1, such that options that scored higher on the intensity and brightness attribute dimensions were assigned a higher nominal value. We explain the choice triplet generation process in more detail below.

First, we determined the location of the target and competitor options. To do this, we first chose a random constant that was substituted into the following equation:

$$Brightness = -0.5 \cdot Intensity + constant. \tag{4.2}$$

In Figure 4.2, the dark blue line is defined by this equation, with $constant = 0.75$. The constant in Equation 4.2 was always chosen to ensure that the point

Figure 4.2: Illustration of the choice triplet selection process.



where the red and blue lines cross fell within the polygon ($0.35 < constant < 1.15$). On Figure 4.2, this is the center of the polygon where $Intensity = 0.5$ and $Brightness = 0.5$, and the dark blue line is thus the reflection of the red line over a vertical line defined by $Intensity = 0.5$ (the middle dashed line on Figure 4.2).

We then created the two target candidates by selecting two points along the reflected line (the dark blue line), one in the upper half of the polygon (target candidate A), and another in the lower half (target candidate B), such that the distance between the two points had to be at least half the overall length of the dark blue line within the polygon. This was important in order to ensure that one of the target candidates had a high intensity and a low brightness value and vice versa, so that they would be perceived as markedly different by the decision maker (as is required from the target and the competitor in an attraction effect choice scenario). Having created the two target candidates, the next step in creating the choice triplets was to determine the position of the decoy option.

To create the decoy, we first randomly decided which target candidate will be assigned a decoy option (Target A or Target B on Figure 4.2). We then again reflected the dark blue line through a vertical line that crossed the chosen target option (the dashed lines defined by $Intensity = 0.25$ and $Intensity = 0.75$ for Target A and Target B, respectively), which, depending on which target candidate was chosen to be the target, gave us one of the grey lines, which is parallel to the original red line. Naturally, any option along this line that lies to the left of the target will have a lower intensity and brightness value, and therefore will be inferior to it.

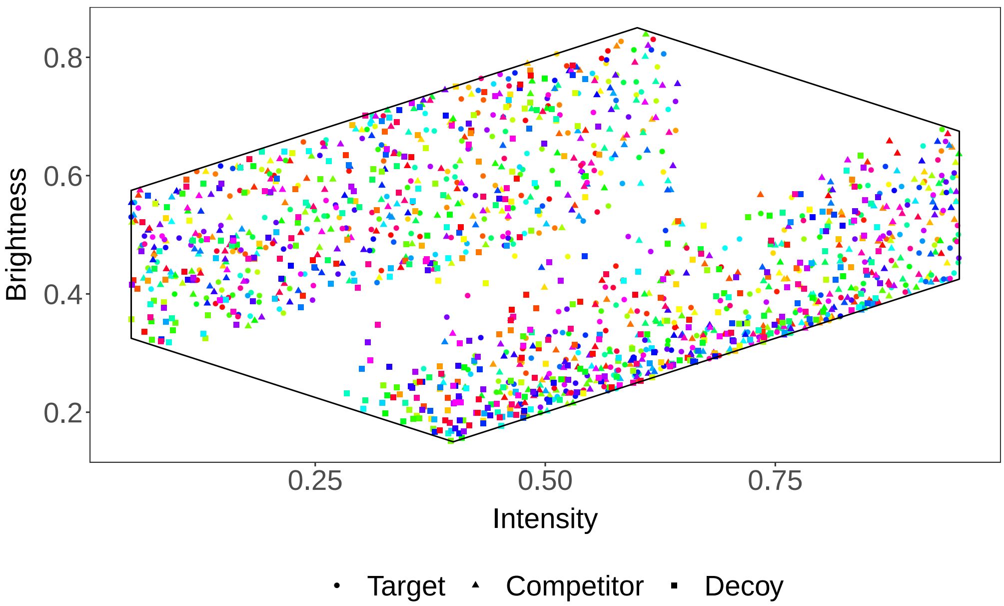
However, all decoys need to satisfy two criteria. First, a decoy needs to be sufficiently far away from the target option, so that the dominance relationship can be easily identified. Perceiving the dominance relationship is especially problematic with naturalistic stimuli (as opposed to a choice scenario with numerical attributes where it can be identified with absolute certainty – provided that the decision maker pays attention to the task). Therefore, we considered this criterion as the most important when creating the decoys. Second, a decoy cannot be too far away from the target, as they need to be somewhat similar to invoke an attraction effect choice situation.

Taking these considerations into account, we decided that the decoy's position should depend on the distance between the target and the competitor (as this can vary to an extent). This avoids situations where the decoy is also dominated by the competitor (this can happen if the target and the competitor are relatively close, while the target and the decoy are relatively far away from each other). Specifically, we decided that the distance between the target and the decoy along the brightness dimension (y axis) should be the 27.5% of the overall difference between the target and the competitor along the same axis. This criterion uniquely defines a point along the grey line (either Decoy A or Decoy B in Figure 4.2).

Once we have decided on the exact brightness and intensity values for the attraction effect choice triplet, the next step was to determine the nominal value of the options. We assigned a nominal value to every point along the red line, starting with £0 and going up to £1000 (in the bottom left and top right corner in Figure 4.2, respectively, where the line crosses the polygon boundaries), so that any line parallel to the dark blue line is essentially a contour line, corresponding to a certain nominal value. Using these contour lines, we can assign a nominal value to any point within the polygon.

For example, as the dark blue line crosses the red line at exactly halfway through its length within the polygon, the target and competitor were assigned a value of £500. Then, once we calculated the position of the decoy, we were able to define the relevant contour line (light blue line in Figure 4.2), which corresponds to a nominal value of £328.

Figure 4.3: The 500 choice triplets that were used in the experiment.



Using the method described above, we generated 500 choice triplets with target, competitor and decoy, as shown on Figure 4.3, where each option is defined by its brightness, intensity and nominal value. In the pictorial condition, participants were presented coloured teapots, where the colour of each teapot was defined by the corresponding intensity and brightness values, and we instructed them to select the teapot with the highest nominal value. We hypothesised that using integer numbers will make comparisons cognitively less demanding in the numerical condition, therefore we transformed the raw brightness and intensity values by subtracting the minimum and multiplying them by 200. The raw intensity and brightness values fell between 0.05–0.95 and 0.15–0.85, and the new range of intensity and brightness values fell between 0–180 and 0–140, respectively.

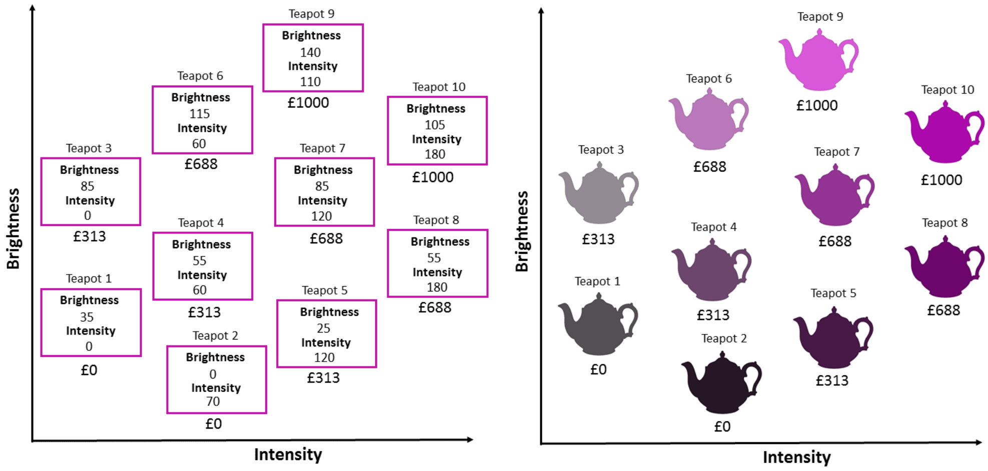
4.2.2.2 Experimental procedure

The experiment consisted of a numerical and pictorial condition, each with two stages: a learning and a choice stage, the latter of which was the main task. All participants were required to complete both conditions, and the order of the two conditions was determined randomly.

The learning stage served to "teach" participants how to infer the nominal value of the options, based on the numeric attributes/colour of the teapots. While a one-dimensional valuation rule can be fairly intuitive (e.g., the brighter or more intense the colour, the better) and thus easy to learn, when it comes to a two-dimensional learning rule, the interactions between the two integral dimensions can complicate the valuation process. For this reason, the choice stage in each condition was only accessible upon passing the corresponding learning stage. Before the learning stage, participants were provided with a "valuation map", which served to explain how the nominal value depended on the attribute values (see Figure 4.4). The example stimuli on the valuation maps were derived from four equally spaced contour lines, like the blue lines on Figure 4.2.

In the numerical condition, we invited participants to try to infer the trade-off between the two attribute dimensions before starting the learning stage. The true underlying trade-off is reflected by the red line on Figure 4.2: a two

Figure 4.4: Value maps for both choice tasks in the experiment.

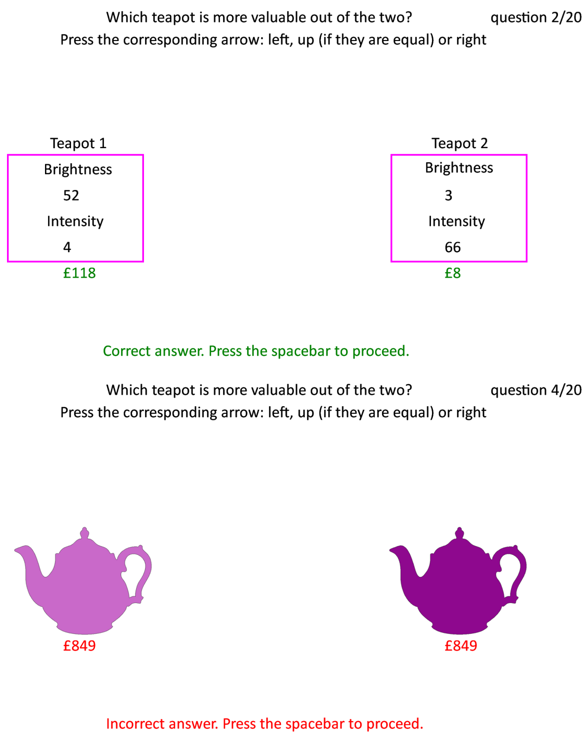


unit change in intensity is equal to a one unit change in brightness.

Each learning stage consisted of 20 questions, where participants had to guess the relative nominal values of the two displayed options, and use the keyboard to indicate which option was worth more than the other (left/right arrow), or whether they were equally valuable (up arrow), as shown on Figure 4.5. Given that indifference between the target and competitor is crucial in invoking an attraction effect choice situation, we wanted to make sure that participants were able to recognise that two markedly different options can have a similar nominal value. After the keypress, participants were given feedback about whether the answer was correct, and were shown the actual nominal value of the displayed options, to facilitate learning. In order to pass the learning stage and proceed to the subsequent choice stage, participants had to get at least 75% of the questions right (at least 15 questions out of 20). However, participants could attempt to complete the learning stage as many times as they wished, and after every failed session, they were encouraged to consult the relevant value map once more before trying again.

To generate the stimuli for the learning stage, we used the 500 choice triplets to create four types of practice questions, based on the value difference

Figure 4.5: Example practice trial in the numerical and pictorial condition.

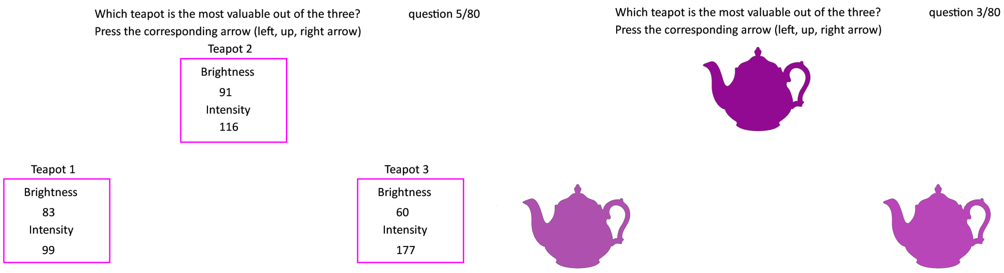


between the two options: equal (where the two options were of equal value), difficult (£100–£150 value difference), moderate (£150–£250 value difference), and easy (value difference higher than £250). Using the 1500 choice options generated for the 500 choice triplets, we created 400 questions for the learning stage, 100 questions for each difficulty level. Each learning stage comprised of 20 questions, with 5 randomly chosen questions from the set of pre-generated questions for each difficulty level.

Once the participant had passed the learning stage, the choice stage began. Participants were presented with 80 choice triplets, and their task was to select the option with the highest nominal value. Out of the 80 questions, 12 were catch trials that served to gauge the attention of participants. On these trials, there was a clearly dominating option out of the three displayed options.

To create stimuli for the catch trials, we again used the pre-generated 1500 options that comprised the 500 choice triplets, by first choosing an option with a nominal value of at least £300, and then two inferior options that were at least worth £100 less than the first, high-value option. Using this method, we created 75 catch choice triplets overall. For each choice stage, the 12 catch trials were randomly chosen from the overall 75 triplets. The rest of the trials in the choice stage were attraction effect choice triplets, and were randomly selected from the 500 choice triplets (see Figure 4.6). During each trial in the learning and choice stage, the presentation order of the options was always randomised. Between the two conditions (pictorial/numerical), participants could take a break for as long as they liked.

Figure 4.6: Example choice triplets (from left to right: DTC and DCT).



To obtain a power level of 80%, we calculated that a sample size of 100 would be suitable, assuming an effect size of 0.25 (Cohen's d). All participants were recruited through the Warwick SONA System. We obtained ethics approval from The University of Warwick's Humanities and Social Sciences Research Ethics Committee (reference number: 50/17-18). The study was advertised as a decision-making study, and participants were given a £4 show-up fee, and were told that they could earn £0.5 for every block of 20 correct questions, therefore, the highest possible earning in this experiment was £8 (as there were 160 questions overall). Given that completing the study involved 160 choice trials (and the two practice sessions), incentive-compatibility was important to ensure that participants stayed motivated during the choice stages. The experiment automatically terminated after 50 minutes if the participant had not finished by then.

4.2.2.3 Exclusion criteria

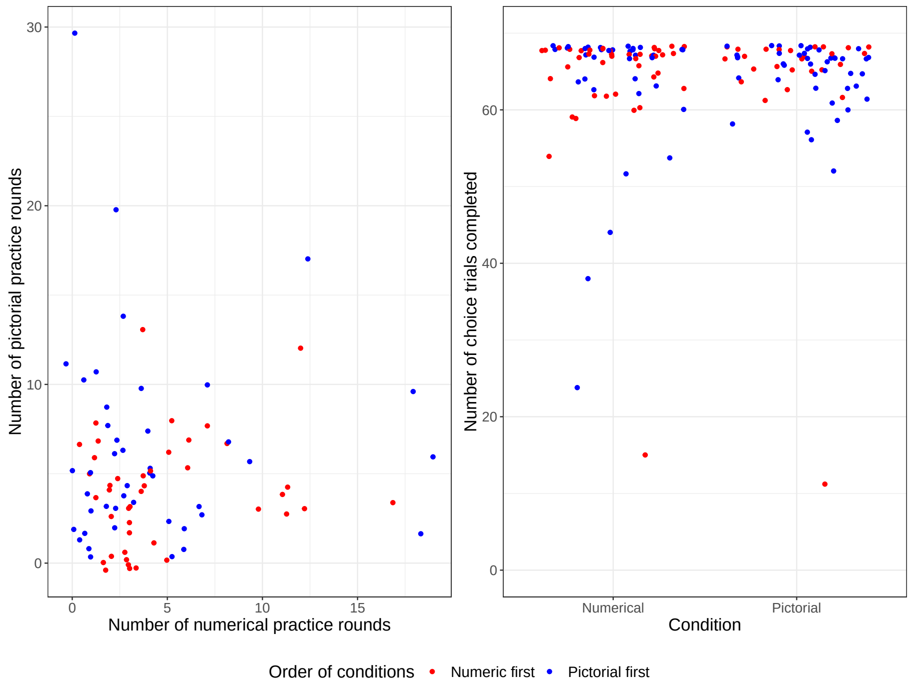
To ensure that we only include participants who took the task sufficiently seriously, we excluded choice blocks where the accuracy on the catch trials was 2.5 standard deviations below the average accuracy for that type of choice task (pictorial/numerical). In addition, we excluded choice blocks that fell into the lowest 2.5% of the entropy distribution, and the upper and lower 2.5% of the autocorrelation distribution across all participants, based on the pattern of their responses. Finally, we excluded trials that fell into the fastest 2.5% of the reaction time distribution, and trials where the subject selected the decoy option. The study design, exclusion criteria and all the analyses were planned and registered before we collected any choice data (see Appendix C for the pre-registration).

4.2.3 Results

Ideally, if all of the 100 participants had completed both conditions, we would have data from 200 choice blocks (100 pictorial and 100 numerical each). However, the learning stages turned out to be more challenging for participants than we originally intended, and 14 people did not manage to pass the first learning stage in 50 minutes (or gave up earlier), and consequently we did not manage to collect any choice data from these participants. After applying all exclusion criteria, we were left with choice data from 86 people, who completed 68 pictorial and 76 numerical choice blocks. Out of the 86 people, 61 completed both the pictorial and numeric conditions within the 50 minutes provided.

The left panel on Figure 4.7 shows the number of attempts it took for participants to pass the numerical and pictorial learning stage (each attempt consisted of 20 questions, as explained above), by the order of the conditions (numerical/pictorial first, therefore, there are 86 dots, each of which is a participant). For the majority of participants, it took fewer than 10 attempts to

Figure 4.7: Number of practice and choice trials by condition and condition order.

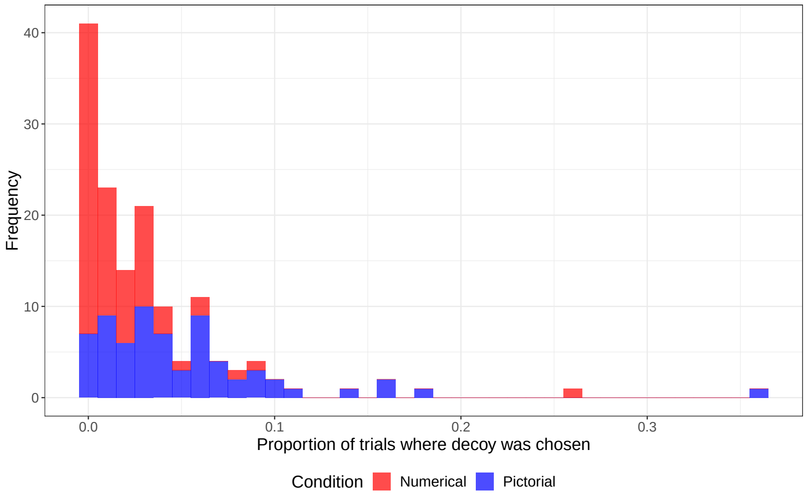


pass the learning stages, and there is considerable individual heterogeneity in the number of attempts it took to complete the learning stage in the two conditions.

The right panel in Figure 4.7 shows the number of completed choice trials by condition and condition order (here each dot refers to a participant's condition, there are 144 of these overall), demonstrating that the vast majority of participants have managed to finish at least one of the choice stages before the experiment was terminated. We can also see that the choice stages were more likely to be interrupted in the numerical condition than in the pictorial condition, because choice trials involving numerical choice options typically took longer.

Figure 4.8 shows the distribution of the proportion of trials on which the decoy was chosen, by condition. It can be seen that participants generally performed well in identifying dominated options (in 94% of the 144 conditions, the proportion of the trials where the decoy was chosen was below 10%). In addition, the decoy was never chosen in 45% of the numerical and 10% of the pictorial conditions, indicating that participants found it much easier to identify

Figure 4.8: Distribution of the proportion of trials on which the decoy was chosen by condition.



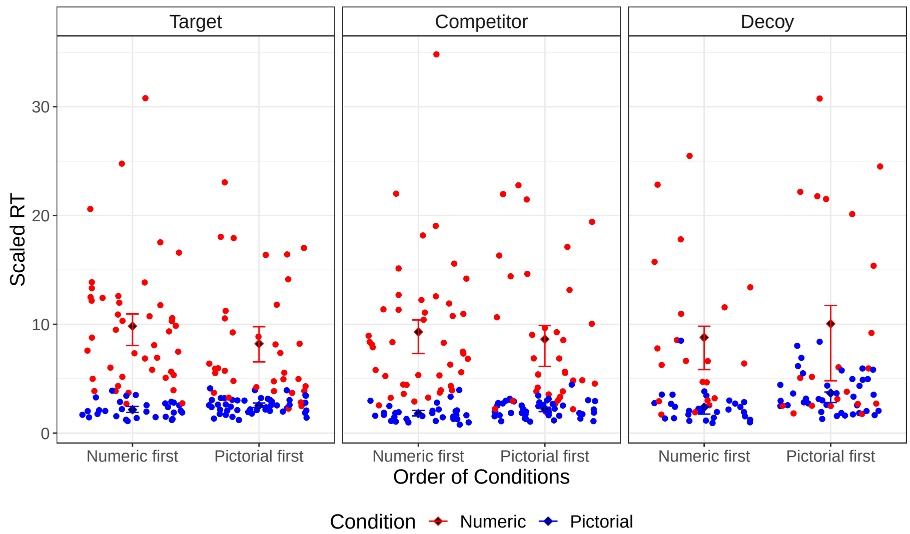
the decoy in the numerical condition. This is not surprising, given that in the numerical version, the choice process requires the sequential comparison of the attribute values, during which the decoy option is more likely to be identified with certainty and thus avoided, whereas in the pictorial version people are more likely to make a quicker, intuitive decision, which might result in an error.

To demonstrate this, Figure 4.9 shows the median of the reaction times (measured in seconds, scaled by subject) by condition, chosen item, and subject. As expected, participants took substantially longer to make a decision in the numerical condition. Interestingly, we do not see substantial differences in reaction times by the chosen item.

In addition, there were no differences in the performance on the catch trials between the numerical and pictorial conditions (paired Wilcoxon Signed-Rank Test using data from the 58 participants who completed both conditions, p = .482), showing that participants were equally good at spotting and choosing a clearly dominating option in the two types of choice tasks.

As specified in the pre-registration, to investigate how the attraction effect depends on the presentation format of the stimuli, we first tested whether the order of the conditions modulated the strength of the attraction effect. We did

Figure 4.9: Distribution of the median scaled reaction times of each participant by condition and chosen item (target, competitor, decoy). Reaction times were first scaled by subject, then the median was calculated for each subject, condition, and chosen item. Black points and corresponding error bars represent bootstrapped 95% CIs of the means of these medians, weighted by the number of trials.



not expect the effect to depend on the order. For each participant and condition, we calculated the attraction effect as the proportion of all trials on which the target was chosen, after excluding trials where the decoy was chosen. The left panel on Figure 4.10 shows the distribution of the proportion of trials on which the target was chosen. We can see that while the order of the conditions does not affect the strength of the attraction effect in the pictorial choice task, there is a pronounced increase in the tendency to choose the target in the numerical condition if it follows the pictorial choice task. The right panel of Figure 4.10 echoes the same pattern, showing the strength of the effect in the two choice tasks for the subset of participants who completed both types of choice tasks (of which there were 58).

Table 4.1 shows the results from a mixed-effects logistic regression, where we explore how the strength of the effect varies by condition and condition order. The results again indicate that the strength of the attraction effect in

Figure 4.10: Proportion of trials on which the target was chosen by condition.

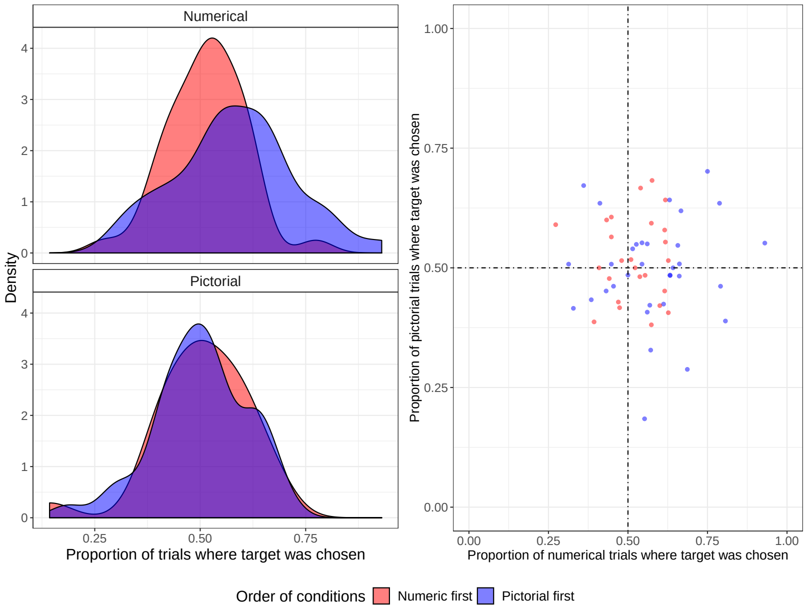


Table 4.1: Odds-ratios from a logistic regression (weighted by the number of trials). 95% CIs are in brackets.

	Dependent variable:
	Target chosen proportion
Condition:Numerical	1.001 (0.876, 1.145)
Order:Pictorial first	0.946 (0.789, 1.137)
Condition:Numerical·Order:Pictorial first	1.405*** (1.173, 1.684)
Constant	1.059 (0.919, 1.219)
Observations	144
Log Likelihood	−508.063
Akaike Inf. Crit.	1,026.126
Bayesian Inf. Crit.	1,040.975
Note:	*p<0.1; **p<0.05; ***p<0.01

the numerical condition is much more pronounced when the first condition was pictorial. Specifically, switching the order of the two conditions (from numerical first to pictorial first) results in a 40%, 95% CI [17%–68%], increase in the odds of choosing the target in the numerical condition.

Since the order of the conditions unexpectedly modulated the strength of the effect, we only used every participant's first condition for our final test of the attraction effect, as specified in our pre-registration. As expected from the visual inspection of Figure 4.10, using Welch's t-test, we found no difference between the proportion of trials where the target was chosen in the pictorial (M = 0.50, 95% CI [0.47, 0.53], N = 41) or numerical (M = 0.51, 95% CI [0.49, 0.54], N = 42) condition, $t(78.87) = 0.65$, $p = .52$.

4.2.4 Discussion

To summarise, our results are somewhat mixed. On the one hand, we have not found evidence for the attraction effect in the pictorial condition, which was expected. On the other hand, we found that the strength of the attraction effect strongly depended on the order of the conditions in the numerical condition, with a markedly higher likelihood of the target being chosen if the numerical condition followed the pictorial condition.

We originally predicted to see the attraction effect in the numerical condition, regardless of the order of the conditions. The choice pattern we found surprised us, for several reasons. First, decades of decision-making research suggests that the attraction effect is a reliable phenomenon when attribute values are displayed numerically, yet in our experiment participants were indifferent between the target and competitor when the numerical condition came first. In addition, it is not immediately obvious why the order of the conditions would have such a strong effect on susceptibility to the attraction effect. We can only speculate about the reasons behind these puzzling results.

One obvious difference between our experimental task and the standard attraction effect choice paradigm is that in our task, participants had to learn the underlying trade-off between the attribute values, as opposed to relying on a more intuitive, preferential valuation rule. This is problematic for two reasons. First, it might mean that our artificial valuation rule was conceptually different from a standard preferential choice task (where valuation is more intuitive), as the choice process involved consulting a previously learnt rule, which is unlikely to resemble a preferential choice process. A second, related issue is that in the practice stages, we trained participants to be experts in inferring the nominal value of the options, while it has been previously shown that the attraction effect tends to be stronger when the decision maker is unfamiliar with the choice domain, and weakens with expertise (e.g., Huber et al., 2014). This is supported by the fact that increasing the number of practice trials slightly reduces the attraction effect, as one additional practice trial decreases the odds of choosing the target by 5.8%, 95% CI [3%, 10.9%] (see regression results in Appendix C.1).

However, even if these concerns have some validity, they do not offer a comprehensive explanation for our results. Specifically, they cannot accommodate the fact that we found a strong effect of condition order. Our results show a strong attraction effect in the numerical condition when it follows the pictorial condition, and no attraction effect when the numerical condition comes first.

One tempting explanation for such a strong temporal effect could be that cognitive fatigue affected participants' decision making in the second half of the experiment. As mentioned before, the learning stages of the experiment proved to be much harder than we expected and intended, illustrated by the fact that 42% of the participants did not manage to complete the experiment within 50 minutes, with 14 people giving up altogether before passing the first learning stage. In addition, the numerical condition required significantly more cognitive effort than the pictorial task, reflected by the much longer reaction times (see Figure 4.9).

It has been previously suggested that depletion of cognitive resources (induced by a Stroop-task that preceded the choice stage) results in a more pronounced attraction effect, due to higher reliance on intuitive decision processes (Pocheptsova, Amir, Dhar, & Baumeister, 2009). This conclusion is supported by the finding that the attraction effect is stronger for those who are more likely to engage in intuitive reasoning (Mao & Oppewal, 2012), although a later replication attempt failed to reproduce the same choice pattern for depleted participants (de Haan & van Veldhuizen, 2015). The link between cognitive fatigue and susceptibility to the attraction effect remains ambiguous in the literature, and we cannot test the role of cognitive fatigue in the context of our choice experiment, as we did not measure fatigue. In addition, the slight negative effect of the number of practice trials on the strength of the attraction effect is not consistent with a fatigue-based explanation.

Previous research has suggested that the attraction effect strengthens when deliberation time increases (Trueblood et al., 2014), although cognitive fatigue, if present, is likely to alter this relationship. While Figure 4.9 shows a very slight tendency towards shorter reaction times when the target is chosen in the numerical condition when it follows the pictorial condition, this difference is not statistically significant (Wilcoxon Signed-Rank Test using data from the 58 participants who completed both conditions, $p = .864$).

If participants got tired towards the end of the experiment, *and* they are more susceptible to the attraction effect in the numerical condition after a prolonged period of cognitive effort, then we would expect that the likelihood of choosing the target increases throughout the choice stage. To test this hypothesis, we explored how the tendency to choose the target changed over time within a condition (numerical/pictorial), depending on the order of the conditions.

To this end, we divided the number of trials in each condition into four equal sized blocks (in their temporal order), and calculated the corresponding target choice proportion for each block. If the overall number of trials were not a multiple of four, then the remainder was allocated to the fourth block.

Figure 4.11: Mean target choice share by condition, condition order and block number. Error bars represent 95% bootstrapped CIs, weighted by the number of trials in each block.

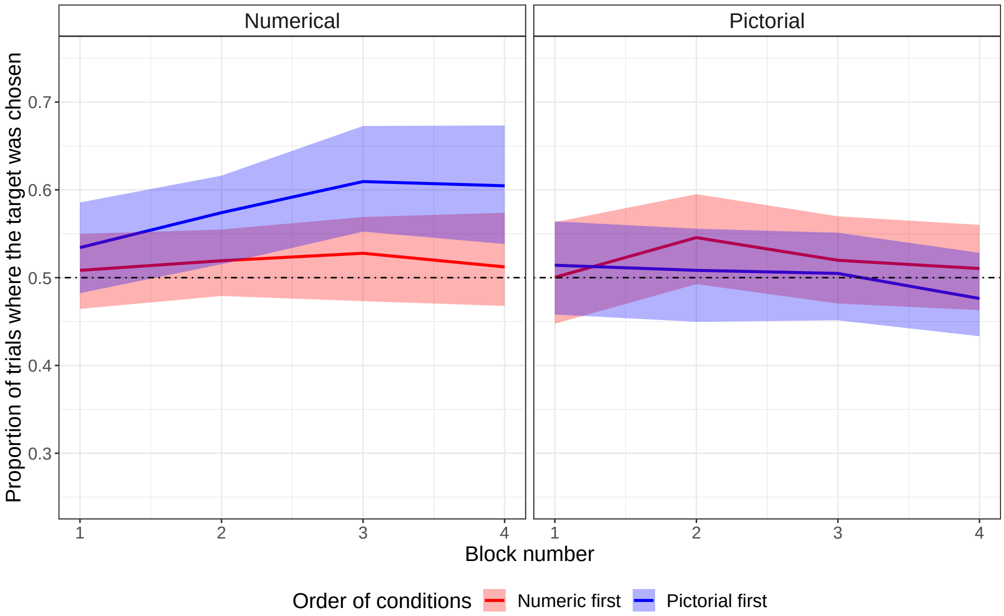


Figure 4.11 shows the average proportion of trials where the target was chosen by condition, condition order, and block number, revealing a slightly increasing temporal tendency to choose the target in the numerical condition if it follows the pictorial condition. Interestingly, the pattern is absent from all other condition-order combinations. While it is likely that our findings are a result of a complex interplay between several factors that we did not predict when we designed the experiment, this temporal pattern suggests that cognitive fatigue could be an appropriate candidate theory for explaining the strong link between the strength of the attraction effect in the numerical condition and condition order in our experiment.

4.3 General Discussion

This study set out to investigate the extent to which the attraction effect depends on the presentation format of the stimuli. Our earlier study described in Chapter 3 had failed to replicate the effect using complex, naturalistic stimuli. We hypothesized that stimulus presentation, and more specifically, the possibility of serial processing of the attribute dimensions might be key in explaining the seemingly elusive nature of this decision bias.

Using insights from decades of psychological research on information processing, we created a simplified, artificial stimulus that can be represented in a pictorial and numerical format, to test if it is the separable nature of the attribute dimensions that gives rise to the attraction effect. We expected to see a strong attraction effect in the numerical condition, and no attraction effect in the pictorial condition.

Somewhat unexpectedly, we found a strong attraction effect in the numerical condition, but only when it followed the pictorial condition. The likelihood of choosing the target showed an upward temporal trend in this numerical condition, consistent with a cognitive fatigue explanation, but our data did not allow us to directly test the link between cognitive fatigue and susceptibility to the attraction effect.

While we cannot offer a comprehensive explanation for these contradictory results, we also recognise that artificially induced preferences (created by our imposed valuation rule) are unlikely to give rise to the same decision processes as more intrinsic preferences, where little or no cognitive effort is required to make a decision. For this reason, we suggest that any test of the attraction effect using perceptual stimuli should aim to use a simple rule to guide choices, as these are more likely to resemble real-life preferential decisions, and thus are better suited to probe the attraction effect under different stimuli formats.

A more comprehensive investigation of the link between the strength of the attraction effect and stimulus presentation format could explore how the separability of the attribute dimensions alter this relationship. For example, a purely numerical condition could be compared to a perceptual version, with spatially separable attribute dimensions (e.g., width and height of rectangles, Trueblood et al., 2013). Then, in another condition, the presentation of the pictorial stimuli could be spatially non-separable, preventing the serial processing of the attribute dimensions. The final condition could test the strength of the effect with integral dimensions (e.g., intensity and brightness), but with no imposed valuation rule. Differences in the strength of the attraction effect between these conditions could provide us with important insights about the cognitive processes underlying this decision bias.

Chapter 5

Exploring the link between football and domestic abuse – Evidence from the West Midlands

5.1 Introduction

Domestic abuse is a complex phenomenon affecting people from all walks of life. It is increasingly recognised as a major public policy concern in many countries, including the UK (Prpic & Rosamund, 2018). While anyone can become a victim of domestic abuse, women are disproportionately affected, with more than 25% of women, and 15% of men in England and Wales reported to have experienced some form of domestic abuse since the age of 16 (Office for National Statistics, 2018).

In the UK, the definition of domestic abuse has changed over time, owing to the increasing recognition of its multifaceted nature. The initial cross-government definition was the following: "any incident of threatening behaviour, violence or abuse [psychological, physical, sexual, financial or emotional] between adults who are or have been intimate partners or family members, regardless of gender or sexuality" (Home Office, 2012). This definition was subsequently updated in 2012 to include the concept of controlling and coercive behaviour and recognise domestic abuse as a pattern of interactions (as opposed to a one-off incident). The preferred terminology has also been changed from "domestic violence" to "domestic abuse", to reflect that the new definition encompasses a wider range of behaviours.

Domestic abuse has substantial mental health implications, with an estimated three-quarter of survivors experiencing posttraumatic stress disorder symptoms. In addition, they and are significantly more likely to report feelings of anxiety and depression, compared to the general population (Ferrari et al., 2014). The long-lasting impact of domestic abuse is not limited to the direct target of abuse. Witnessing domestic abuse at home can have severe developmental impacts on children, including an increased likelihood of experiencing mental and physical health problems and difficulties in interpersonal relationships in later life, worse educational attainment, and engagement in criminal behaviours

(Callaghan, Alexander, Sixsmith, & Fellin, 2018).

As the cross-government definition reflects, one characteristic of domestic abuse that differentiates it from other types of violent crime is its repeated nature. It is estimated that on average, victims live in the abusive relationship for 2.7 years, experiencing approximately 50 cases of abuse before getting effective help (*Ending Domestic Abuse - Getting it right first time*, 2015). The most reliable statistics on domestic abuse in the UK is the Crime Survey for England and Wales (CSEW; Office for National Statistics, 2018), a victimisation survey that includes a self-completion module on domestic abuse. According to the CSEW, only 17% of those respondents who experienced domestic abuse between April, 2017 and March, 2018 reported it to the police (Office for National Statistics, 2018). This extremely high level of underreporting is another characteristic that differentiates domestic abuse from other types of crime.

In the most extreme cases, domestic abuse can culminate in domestic homicide. In the period between April, 2017 and March, 2018, 70 people in England and Wales were killed by their current or former partner, and 90% of these victims were women (Office for National Statistics, 2019), demonstrating that domestic abuse is a fundamentally gendered phenomenon. While the pervasive problem of underreporting poses a significant obstacle to deriving reliable estimates of the true extent of the problem, the economic cost of domestic abuse in England and Wales between April, 2016 and March, 2017 was estimated to be as high as £66 billion (R. Oliver, Alexander, Roe, & Wlasny, 2019). The largest component of this cost is represented by the physical and emotional consequences of abuse, reflected in the reduced expected quality of life for survivors. In addition, lost economic output resulting from missed workdays and reduced productivity, as well as costs to the health care system also significantly contribute to the overall figure.

In light of these costs, substantial research efforts have been devoted to the identification of common precipitating factors of domestic abuse. While causal inference is impossible due to the complexity and multifaceted nature of the phenomenon, identifying the factors that make people more likely to become perpetrators is crucial for designing effective interventions. Studies exploring the risk factors of domestic abuse perpetration have found that young age, substance abuse, low socio-economic status, and low educational attainment are all associated with an increased risk of becoming a perpetrator (e.g., Capaldi, Knoble, Shortt, & Kim, 2012; Costa et al., 2015; Yakubovich et al., 2018). In addition, ample quantitative evidence shows that economic deprivation is a significant predictor of domestic abuse victimisation, especially for women (e.g., Walby & Allen, 2004; Khalifeh, Hargreaves, Howard, & Birdthistle, 2013; Towers, 2015; Fahmy & Williamson, 2018), suggesting that the lack of economic independence is a major factor preventing women escaping from the cycle of abuse (Walby, Towers, & Francis, 2016).

In the wake of the 2018 FIFA World Cup, a powerful, chilling poster campaign named "The Not-So-Beautiful-Game" was launched by the National Centre for Domestic Violence, to raise awareness of the link between the upcoming World Cup and the expected increases in the number of reported domestic abuse incidents (National Center for Domestic Violence, 2018). The poster featured a picture of a woman's face with St George's cross painted over it with blood, recreating the English flag. The text read: "If England get beaten, so will she. Domestic violence increases 26% when England play, 38% if they lose" (see Figure 5.1). The campaign has proved to be highly successful, and had been widely discussed in the British media before the 2018 World Cup. Many England football fans were surprised and disturbed by the association of their beloved game with domestic abuse, in spite of the well-documented association between team contact sports and violence.

Figure 5.1: “The Not-So-Beautiful-Game” campaign by the National Centre for Domestic Violence. https://www.ncdv.org.uk/the-not-so-beautiful-game/. Copyright 2019 by NCDV. Adapted with permission.

5.1.1 Making the link between sport and domestic abuse in England

The link between sport and violence has long been the focus of academic research. While sporting activities have a range of widely recognised societal benefits, violence is undoubtedly an inherent part of many team contact sports, either within, or outside the rules of the game (Swallow, 2017). In nineteenth-century England, modern team sports (e.g., rugby, cricket, football) were seen as “manly sports”, and were regarded as excellent ways to masculinize and prepare young (upper- and middle-class) boys for their future careers, through the principles of endurance, loyalty, and respect for authority. At the same time, girls were explicitly deterred from practising these sports, corroborating the notion that men are “a breed apart” (Kidd, 2013). This resulted in socially accepted gender differences in participation rates in these sports that still persist. Today, televised sports provide easy access to the sports fan experience for everyone. However, research in sport sociology suggest that there still exist significant differences in the role and nature of sports fandom between genders (Sabo, Gray, & Moore, 2000).

Traditionally masculine values, such as strength, toughness, aggression, power and dominance are embedded in sports. If these values can be internalised through spectatorship, then it is a natural concern that they may manifest in the real life actions of some fans in the form of violence (Williams & Neville, 2014). Previous qualitative research has indeed suggested that televised contact sports can serve as vehicle for the male sports fan to redefine, and express his masculinity in a way that allows dominance, control, and can ultimately manifest in the perpetration of violent acts, including domestic abuse (Sabo et al., 2000; Swallow, 2017), given susceptibility to such behaviours. Below we review the quantitative evidence for this relationship.

5.1.2 Evidence for the link between sport and domestic abuse

There exist numerous investigations of the link between sport and domestic abuse. In the US, most studies have focused on the National Football League (NFL). White, Katz, and Scarborough (1992) tested whether the number of violent assaults (measured by the number of emergency room admissions) in northern Virginia were affected by Washington Redskins football games between 1988–89, depending on the outcome of the match. Controlling for day of the week, month, year, as well as public holidays, they found that the number of women admitted to emergency increased significantly on the day following a victory of the Redskins. However, a limitation of this study is that they could not identify specifically domestic abuse related assaults.

In a later exploratory study, Sachs and Chu (2000) analysed the relationship between NFL games and the number of dispatches to domestic disputes by the Los Angeles County Sheriff Department between 1993 and 1995. They compared the percentage increase in the number of dispatches from Wednesdays (no games played) to Sundays (regular game day), and tested whether this increase was significantly higher on football, playoff, and Super Bowl weeks, compared to no-football reference weeks. Surprisingly, they observed the opposite pattern in the two football seasons, with a general increase on playoff and Super Bowl weeks in the 1993–94 season, but a decrease in the following season, albeit both nonsignificant compared to baseline (no-football weeks). They speculate that this difference might stem from the fact that while 1993–94 was a winning season for the local team, a year later they made plans to move to a different city, which might have affected fans' engagement with the games.

In a more extensive study, Gantz, Bradley, and Wang (2006) tested whether the number of domestic violence dispatches increased on NFL game days, using data from 15 cities with NFL franchises between 1996 and 2002, controlling for day of the week, month, year, and public holidays. They found a modest increase in the number of domestic violence dispatches on days when the local team played an NFL game.

Oths and Robertson (2007) explored the temporal pattern of crisis calls made to a domestic violence shelter between 1997 and 1999. They found that on average, there are *fewer* calls on the day of the Super Bowl, or on other holidays (including Labor Day, Halloween, Memorial Day, Fourth of July, Easter, Christmas, New Year's Eve, and Thanksgiving), compared to other days during the year. In contrast, their data suggest that the number of calls increase during school holidays, and they concluded that the decision to leave an abusive relationship is likely to be guided by the timing of the school vacation, to make it the least disruptive for the children involved.

However, their data show markedly different temporal patterns compared to police data: for example, they report the number of calls to be the lowest during the weekend, whereas police data consistently show that the number of reported domestic abuse cases increase significantly during the weekend (e.g., Liam Bannon, 2016, Scottish Government, 2018). This stark discrepancy might be due to the fact that abuse is more likely to occur when the victim and the perpetrator are together (during the weekend), but the victim is more likely to seek shelter options when the perpetrator is not around (during the week).

Card and Dahl (2011) presented the most extensive study on the link between reported number of domestic abuse incidents and NFL games up to date. Using data on domestic violence reports to the police in 750 city and county police agencies from the period between 1995 and 2006, they found that an unexpected loss of the local football team resulted in a 10% increase in the rate of reported male to female intimate partner violence (IPV). After controlling for the expectations of the fans (through the pre-match betting odds), the result of any match can be considered random, which allowed them to make a causal inference about the effect of NFL games on the number of reported domestic abuse incidents.

In Scotland, two studies have investigated the association between Celtic and Rangers matches (commonly known as “Old Firm” fixtures) and domestic abuse. The famous rivalry between the two Glasgow-based teams has deep political and sociological roots, and their fixtures are known for notorious fan violence (Williams & Neville, 2014). Using data from Strathclyde Police, Williams, Neville, House, and Donnelly (2013) have investigated whether the number of reported domestic abuse incidents increases on days when there is an Old Firm or Scotland International fixture in Glasgow, using data from seven days after the match as a baseline measure. They found that the number of reported domestic abuse cases was significantly higher on Old Firm match days, compared to both Scotland International match days, and non-Old Firm fixture days.

In a later study, Dickson, Jennings, and Koop (2016) have investigated whether the reference point behaviour observed in the context of the NFL in the US (reported by Card & Dahl, 2011) are also present in the behaviour of Celtic and Rangers fans. While their findings demonstrate a 28–41% increase in the number of reported domestic abuse incidents on days when there is an Old Firm match, they found no evidence for an additional effect of unexpected losses. This difference highlights the importance of the cultural context when analysing the link between sport and domestic abuse. However, as Rangers and Celtic fans are not geographically segregated within Glasgow, they had to assume that the other team's upset win does not increase the likelihood that their fans will become perpetrators, which is a strong, untested assumption.

In England, most studies have focused on the link between national football tournaments (such as the World Cup) and domestic abuse. Football's history is inextricably linked to England, and it is by far the most popular sport in the country (Parry, Jones, & Wann, 2014), with the 2018 World Cup attracting a record number of 44.5 million viewers (BBC Sport, 2018). England's participation in these tournaments are times of heightened patriotic emotions, and a strengthened sense of "Englishness", fuelled by media narratives that often use war references, and a "us vs. them" rhetoric to generate and represent an English national identity (Vincent & Harris, 2014). It can be argued that the observation that football fandom can serve as an important tool for masculinity construction is especially pertinent in the context of England's participation in national football tournaments, owing to the popularity of the sport in the country, the associated media attention, and the resulting heightened sense of national consciousness.

In line with this hypothesis, studies using data on emergency department attendances have identified substantial increases in the number of assault attendances on days when England plays in the World Cup (Quigg, Hughes, & Bellis, 2013; Bellis et al., 2012). One of the earliest examinations of the link between football and domestic abuse by Brimicombe and Cafe (2012) used daily data from 33 out of 39 police forces in England from the period of June–July in 2009 and 2010 (World Cup tournament year). They tested whether the reported number of domestic abuse cases increased significantly on days when the England national football team won, lost, or drew, compared to the same days in 2009, and other, non-match days during the tournament in 2010. The study found that rates of reported domestic abuse increased significantly when England lost or won (about 33–35%), but did not change on days when they drew.

A more comprehensive investigation, using daily counts of domestic abuse in Lancashire from the 2002, 2006 and 2010 World Cup, found a 38% increase in the number of reported domestic violence cases when the England team lost, and a 26% increase when they won or drew (Kirby, Francis, & O'Flaherty, 2014). These estimates had been widely discussed in the British media before the 2018 World Cup, and the figures were also quoted on the posters in the Not-So Beautiful Game Campaign.

While domestic abuse is predominantly understood as a pattern of ongoing behaviour involving a series of occurrences, rather than a one-off incident triggered by football (Brooks-Hay & Lombard, 2018), these studies, and other qualitative investigations (e.g., Swallow, 2017) nevertheless suggest that national football tournaments can create an environment for abusers that is conducive to domestic abuse. In the context of England, most studies have hypothesized alcohol to be a significant contributing factor, but we are not aware of any quantitative investigation exploring the role of alcohol in the link between football and domestic abuse.

5.1.3 Theorizing the role of alcohol in the link between football and domestic abuse

Qualitative investigations suggest that alcohol can be a significant factor in the link between football and domestic abuse. Alcohol has a strong association with domestic abuse (Peralta, Tuttle, & Steele, 2010), those with alcohol-problems are more likely to be perpetrators and, when alcohol is involved, there is evidence that the violence might result in more serious injuries. However, it is generally understood that the role of alcohol should be considered in the context of a range of social, biological and psychological factors, and that alcohol is not the direct cause of domestic abuse (Javaid, 2015; Peralta et al., 2010). One explanation for the co-occurrence of domestic abuse and alcohol is that, for some men, drinking and violence plays an instrumental role in the construction and expression of masculinity, especially when the problem of masculine deficiency is present (e.g., by unemployment, Peralta et al., 2010). It has also been suggested that some perpetrators use alcohol to deflect responsibility for their actions, using alcohol as a "shield" that protects them from being seen as a violent abuser (Javaid, 2015).

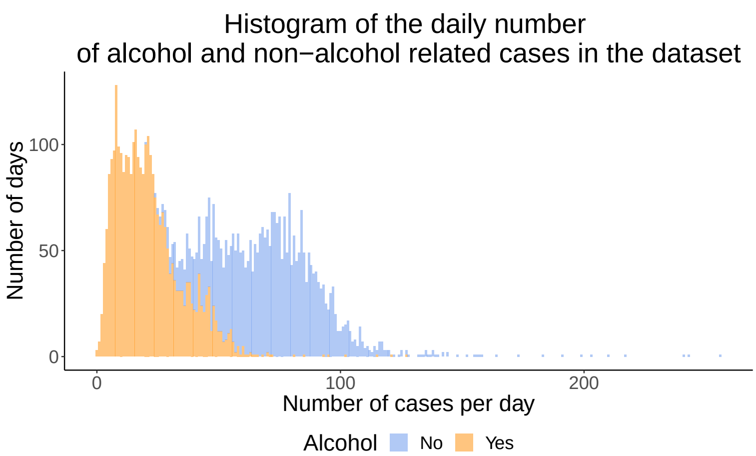
In the US, the relationship between unexpected NFL losses and IPV did not depend on alcohol-involvement in the abuse case (Card & Dahl, 2011), while England-based quantitative studies did not look at the role of alcohol in particular. Given the strong association between drinking culture and football in England (Dixon, 2014), a relationship continuously reinforced by the marketing practices of the alcohol industry (Gornall, 2014), we hypothesize that alcohol play an important role in the relationship between national football tournaments and domestic abuse in England.

To explore this hypothesis, we test if the daily number of reported domestic abuse cases recorded by the West Midlands Police between 2010 and 2018 increase on days when the England national team plays in the World Cup or the European Championship, and whether the effect, if any, depends on alcohol-involvement in the case or the result of the match.

5.2 Data

Our dataset comprises all crimes and specific types of incidents (such as domestic abuse) that have been reported to the West Midlands Police (WMP) in the period between 2010 and 2018. The WMP is the third largest police force in England (Home Office, 2018), serving an estimated 2.9 million people in 2017 (Office for National Statistics, 2017). The first half of 2017 has been excluded due to missing data. The number of reported domestic abuse cases is the sum of crimes that have a domestic abuse marker, and all domestic abuse incidents. Crimes that have a domestic abuse marker indicate cases of domestic abuse that meet the criteria for notifiable offences in the UK, whereas domestic abuse incidents refer to cases that do not qualify as a crime. For each record in this dataset, we have information about the time and location of the incident or crime, and the gender and age of the offender and victim. We restricted our analyses to cases with one victim and one offender. We can also identify repeat offenders and victims by their unique person identifier. Domestic abuse cases constitute about 31% of all recorded crimes and incidents in the dataset, and about 23% of all domestic abuse cases are alcohol-related. The alcohol marker for each case in our dataset is a dummy variable (Yes/No). Therefore, in our dataset, each day in the relevant time period has two rows: one recording the number of alcohol-related and another recording the number of non-alcohol related cases recorded on that day (Figure 5.2 shows their respective distributions). In the period between 2010 and 2018, the daily rate of non-alcohol related domestic incidents fell between 1.6–3 cases per 100,000 individuals, whereas the daily rate of alcohol-related cases fell between 0.35–1 cases per 100,000 individuals. There were three World Cups (2010, 2014, 2018) and two European Championships (2012, 2016) in the period covered by our dataset. All included tournaments took place in the months of June and July.

Figure 5.2: Histogram of the daily number of cases by alcohol involvement



Previous research has mostly focused on IPV, the largest subcategory of domestic abuse. While IPV is more common than abuse perpetrated by family members (Office for National Statistics, 2018), our dataset does not contain information about the exact relationship between the victim and perpetrator, therefore we cannot separate the two types of abuse, and we will refer to them collectively as "domestic abuse".

While our dataset contains all cases of domestic abuse that have been reported to the West Midlands Police between 2010 and 2018, as mentioned before, the vast majority of domestic abuse incidents in fact never get reported. This substantial reporting bias, and its potential correlation with other contextual factors warrant a careful interpretation of the estimates from any quantitative study investigating domestic abuse, and highlights the importance of utilising a mixed methods approach in exploring the factors precipitating domestic abuse.

5.3 Results

In the following regressions, each observation is a day in the period between 2010 and 2018, and the outcome variable is the number of domestic abuse cases reported to have been perpetrated on that day. To investigate whether national football tournaments affect the number of reported abuse cases, we classify each day in our dataset as either a day on which England won (England win, 8 days), lost (England lost, 8 days) or drew (England draw, 6 days), a day after an England match day (After England, 22 days), any other day during the tournament (Tournament on, 106 days), or any other day during the rest of the year (Non-tournament day, 2867 days). All day-level regressions presented below include Christmas, New Year's Eve, day of the week, month and year controls. Since out outcome variable is a count (the daily number of alcohol and non-alcohol related domestic abuse cases reported to the police), we used a Poisson and negative binomial regression framework to investigate the link between football and domestic abuse. More specifically, the regression model we estimated took the following form:

$$ln(C_i) = \mu_i = \alpha + \beta X_i \tag{5.1}$$

with $C_i \sim Poisson(\lambda_i)$ or where $C_i \sim Negative\ Binomial(\lambda_i \theta)$

where C_i is the observed daily count of domestic abuse cases reported to the police (either alcohol, or non-alcohol related), X_i captures the explanatory variables, most importantly Alcohol (Yes/No), Type of day, and a set of control variables (day of the week, month, year, Christmas and New Year's Eve), λ_i is the daily mean of the number of reported cases, whereas θ is the overdispersion parameter for the negative binomial distribution.

5.3.1 Main results

5.3.1.1 The effect of England football games on reported domestic abuse

Using a series of negative binomial regressions, we first compare various, increasingly complex model specifications to understand the relationship between football, as shown in Table 5.1. In the first regression, the coefficient of Alcohol reflects the base rate of alcohol versus non-alcohol related cases in the dataset: there are 72% fewer alcohol-related cases than non-alcohol related cases. Adding type of day as an explanatory variable to a model with only alcohol and time controls marginally improves the model fit (see column 2), and the results show a 20%, 95% CI [5%–38%] increase in the number of reported domestic abuse cases when the England national football team wins. The comparison between column 2 and 3 reveals that this increase stems from a much more pronounced, 61% 95% CI [24%–110%] increase within the subgroup of alcohol-related domestic abuse cases on days when England wins. Interestingly, we find no evidence for comparable increases in the number of reported domestic abuse cases when

Table 5.1: Number of reported domestic abuse incidents by alcohol involvement and type of day. Each day in the analysed corresponds to two observations in the dataset.

	Dependent variable:			
	Number of reported domestic abuse cases per day			
	(1)	(2)	(3)	(4)
Alcohol	−0.719***	−0.719***	−0.719***	−0.862***
	(0.007)	(0.007)	(0.008)	(0.031)
Tournament on		−0.004	0.014	0.032
		(0.023)	(0.027)	(0.020)
England win		0.205***	−0.037	−0.031
		(0.069)	(0.091)	(0.063)
England draw		0.025	0.048	0.047
		(0.082)	(0.104)	(0.072)
England loss		0.078	−0.013	0.050
		(0.068)	(0.089)	(0.061)
After England		0.097**	0.075	0.086**
		(0.043)	(0.055)	(0.038)
Tournament on:Alcohol			−0.043	−0.083**
			(0.040)	(0.035)
England win:Alcohol			0.610***	0.606***
			(0.135)	(0.101)
England draw:Alcohol			−0.055	−0.034
			(0.165)	(0.129)
England loss:Alcohol			0.223	0.076
			(0.135)	(0.101)
After England:Alcohol			0.051	0.037
			(0.084)	(0.066)
Number of observations	6034	6034	6034	6034
AIC	45,539.500	45,536.770	45,530.360	41,959.280

[a] *$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$*

[b] *Estimates are from a series of negative binomial regressions (based on tests of overdispersion) with year, month, day of week, Christmas, New Year's eve controls; Model 4 further includes interactions between alcohol and all control variables; standard errors in parentheses*

the England national team loses. Less surprising, and more consistent with previous findings is the lack of an increase on England draw days, probably due to the fact that high-stake matches after the group-stage in the tournament cannot result in a draw.

Further interacting alcohol with the rest of the time-specific control variables results in a substantially improved model fit (see column 4), but does not alter the effect of an England win on alcohol-related domestic abuse (61%, 95% CI [32%–96%]). The results also reveal a smaller, 9%, 95% CI [1%–17%] increase in non-alcohol related cases on days following an England match day, potentially the result of a temporal spillover effect from the previous match day. We also see an 8%, 95% CI [2%–14%] decrease in alcohol-related cases during the tournament, but not on England match days, perhaps stemming from heavy drinking being mostly concentrated around England match (and particularly England win) days, and relatively lower alcohol consumption on other days during the tournament.

While this increase is proportionally high, and behaviourally important, it is exceeded by other occasions when people are likely to drink. For example, the increase in the number of alcohol-related domestic abuse cases on Saturdays and during the Christmas period are 75% and 85%, respectively.

5.3.1.2 The effect by offender-victim gender group

To explore the characteristics of this increase, we investigate whether the strength of the effect varies by offender-victim gender subgroup. Previous qualitative research has suggested that the link between football and domestic abuse is a result of violent expression of masculinity (Sabo et al., 2000), where heavy drinking is also often present. If this was the case, we would expect football and alcohol to only affect reported numbers of male-perpetrated domestic abuse.

Table 5.2 shows the results from four negative binomial regressions, one for each offender-victim gender group. These reveal a pronounced increase in the subgroup of male to female abuse (which comprises about 80% of all domestic abuse cases in our data), where the number of reported alcohol-related cases increase by 67%, 95% CI [35%–107%] on England win days. While we see similar tendencies for alcohol-related cases in other gender subgroups on England win days, these coefficients are about half the size of the male to female effect, and are not statistically different from zero. These results can be interpreted in light of the observation that British football fandom is prevalently male-dominated (Parry et al., 2014), and they lend support to the hypothesis that masculinity

Table 5.2: Number of reported domestic abuse incidents by type of day, alcohol involvement, and gender of perpetrator and victim

	Dependent variable:			
	Number of reported domestic abuse cases per day			
	Male to Male	Male to Female	Female to Female	Female to Male
	(1)	(2)	(3)	(4)
Alcohol	-0.825^{***}	-0.870^{***}	-0.808^{***}	-0.858^{***}
	(0.101)	(0.034)	(0.133)	(0.080)
Tournament on	0.005	0.038^{*}	0.053	-0.048
	(0.054)	(0.021)	(0.062)	(0.045)
England win	-0.068	-0.022	0.019	-0.147
	(0.165)	(0.066)	(0.193)	(0.135)
England draw	0.080	0.038	0.043	0.107
	(0.194)	(0.076)	(0.225)	(0.169)
England loss	-0.063	0.065	-0.036	0.117
	(0.162)	(0.064)	(0.171)	(0.136)
After England	-0.036	0.093^{**}	0.152^{*}	0.025
	(0.103)	(0.040)	(0.114)	(0.082)
Alcohol:Tournament on	-0.181^{*}	-0.077^{**}	-0.018	-0.215^{*}
	(0.106)	(0.038)	(0.137)	(0.084)
Alcohol:England win	0.334	0.674^{***}	0.360	0.472
	(0.285)	(0.108)	(0.358)	(0.231)
Alcohol:England draw	-0.282	0.031	0.071	-0.580
	(0.411)	(0.138)	(0.629)	(0.313)
Alcohol:England loss	0.286	0.028	0.328	-0.088
	(0.279)	(0.111)	(0.356)	(0.231)
Alcohol:After England	0.209	0.052	-0.111	-0.040
	(0.185)	(0.071)	(0.242)	(0.159)
Number of days	3,017	3,017	3,017	3,017

[a] $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

[b] *Estimates are from a series of negative binomial regressions (based on tests of overdispersion) with year, month, day of week, Christmas, New Year's eve controls interacted with alcohol; standard errors in parentheses*

construction and alcohol may be key to the link between football and domestic abuse. However, it is unclear why victory-induced, alcohol-related masculinity construction would culminate in violence only against women.

5.3.1.3 The effect on other criminal behaviours

Our unique dataset further allows us to explore whether England games have similar effects on other types of criminal behaviours. Specifically, we are interested in whether England match days affect the number of reported property-related crimes (including burglary, theft and robbery), public order offences (behaviours that cause offence to the general public), hate crimes (hate incidents and any other racially or religiously aggravated crime), and other violent crimes (excluding cases of domestic abuse). Of particular interest is the effect of football on non-domestic violent crimes, since it is possible that alcohol-fuelled violence that follows an England victory is not limited to family and intimate partner relationships.

Table 5.3 shows the results from a series of negative binomial regressions for different types of criminal behaviours. These reveal that while there is no evidence that England matches affect the number of reported property-related offences, we see an increase in the number of non-alcohol related public order offence cases on tournament days, when England wins, and on days after an England game. Hate incidents with no alcohol involvement also increase when the tournament is on. But most importantly, the effect of an England match on alcohol-related cases extends to other, non-domestic violent offences, resulting in a 55%, 95% [43%–72%] increase on days when England wins, and a smaller increase on days following an England match, the exact same pattern we have seen for domestic abuse. This result highlights that alcohol-related violent behaviour on England win days is not limited to family relationships. Further analysis reveals that the increase in these alcohol-related non-domestic violent crimes also predominantly comes from male to female cases (although male to male and

Table 5.3: Number of reported cases for each crime type, by type of day, and alcohol involvement

	Dependent variable:			
	Number of reported domestic abuse cases per day			
	Property-related	Public Order Offences	Hate incidents	Other violence
	(1)	(2)	(3)	(4)
Alcohol	-0.981^{***}	-0.922^{***}	-0.934^{***}	-0.902^{***}
	(0.065)	(0.080)	(0.115)	(0.040)
Tournament on	0.042	0.096^{**}	0.138^{***}	0.034
	(0.026)	(0.036)	(0.047)	(0.027)
England win	0.052	0.234^{**}	0.073	0.094
	(0.074)	(0.095)	(0.136)	(0.077)
England draw	0.100	−0.065	−0.066	0.035
	(0.085)	(0.128)	(0.168)	(0.092)
England loss	−0.042	0.075	0.011	0.089
	(0.078)	(0.100)	(0.139)	(0.078)
After England	0.052	0.161^{**}	0.141	0.108^{**}
	(0.047)	(0.062)	(0.084)	(0.048)
Alcohol:Tournament on	0.135	-0.197^{**}	-0.215^{*}	−0.009
	(0.080)	(0.101)	(0.141)	(0.051)
Alcohol:England win	0.259	0.020	0.310	0.507^{***}
	(0.219)	(0.256)	(0.359)	(0.132)
Alcohol:England draw	0.060	0.374	0.393	0.360^{*}
	(0.264)	(0.303)	(0.431)	(0.161)
Alcohol:England loss	0.144	0.456^{*}	−0.032	0.018
	(0.226)	(0.228)	(0.393)	(0.138)
Alcohol:After England	0.094	0.127	0.446^{*}	0.053
	(0.144)	(0.158)	(0.211)	(0.088)
Number of days	3,017	3,017	3,017	3,017

[a] *$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$*

[b] *Estimates are from a series of negative binomial regressions (based on tests of overdispersion) with year, month, day of week, Christmas, New Year's eve controls interacted by alcohol; standard errors in parentheses*

Table 5.4: Non-domestic violent cases by gender

	Dependent variable:			
	Number of reported other violent abuse cases per day			
	Male to Male	Male to Female	Female to Female	Female to Male
	(1)	(2)	(3)	(4)
Alcohol	-0.900^{***}	-0.891^{***}	-0.864^{***}	-0.921^{***}
	(0.048)	(0.033)	(0.080)	(0.066)
Tournament on	0.037	0.050^{**}	0.041	0.051
	(0.026)	(0.021)	(0.038)	(0.036)
England win	0.013	0.019	-0.031	0.174
	(0.082)	(0.067)	(0.111)	(0.112)
England draw	0.089	0.012	0.115	0.042
	(0.094)	(0.078)	(0.139)	(0.132)
England loss	0.018	0.028	0.088	0.118
	(0.082)	(0.066)	(0.114)	(0.108)
After England	0.085	0.070	0.181^{**}	0.149^{**}
	(0.050)	(0.042)	(0.071)	(0.067)
Alcohol:Tournament on	-0.027	-0.086^{**}	-0.077	-0.167^{**}
	(0.055)	(0.038)	(0.087)	(0.073)
Alcohol:England win	0.391^{**}	0.613^{***}	0.441^{*}	-0.114
	(0.158)	(0.109)	(0.251)	(0.199)
Alcohol:England draw	0.071	0.102	0.127	-0.337
	(0.192)	(0.137)	(0.361)	(0.254)
Alcohol:England loss	0.296^{*}	0.057	-0.023	0.027
	(0.153)	(0.112)	(0.237)	(0.207)
Alcohol:After England	0.208^{*}	0.053	-0.119	-0.158
	(0.100)	(0.072)	(0.163)	(0.136)
Number of days	3,017	3,017	3,017	3,017

[a] *$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$*

[b] *Estimates are from a series of negative binomial regressions (based on tests of overdispersion) with year, month, day of week, Christmas, New Year's eve controls interacted with alcohol; standard errors in parentheses*

female to male cases also contribute, see Table 5.4). While it is possible that a number of misclassified domestic abuse cases are reflected in this result (e.g., if the victim refuses to admit any relationship to the offender), but even if this was the case, taken together, these findings only strengthen our conclusion that football and alcohol primarily make men more violent, and direct this violence overwhelmingly towards women.

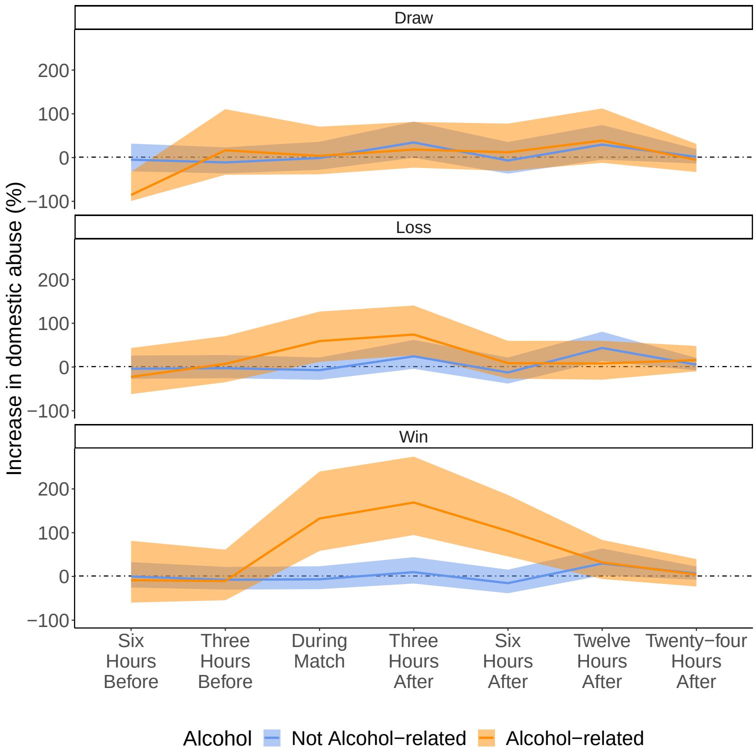
5.3.1.4 Three-hour analysis of the effect

Next, we explore the temporal dynamics of the increase in alcohol-related domestic abuse on England match days in more detail. Our previous results revealed important differences in the effect of football on domestic abuse depending on alcohol-involvement in the case, therefore we run two separate regressions for alcohol and non-alcohol related domestic abuse cases to analyse the temporal pattern of the increase. In our dataset, the end and start time of the incident, as well as the time of reporting is recorded to the minute.

To explore the temporal dynamics of the England win effect, we divided each day in our dataset into eight three-hour periods, the first one starting at 12am, and used these to identify specific time windows around the time of the match. The exact time of the matches vary considerably (the earliest starting at 1pm, and the latest at 11pm). We first identified the three-hour period of the day into which each match falls. If the start and end time of the match did not fall in the same three-hour period, we chose the three-hour period that covers the larger part of the match (e.g., a 2.5 hour long match starting at 7pm will be assigned to the 6–9pm period and not to the 9pm–12am period).

Figure 5.3 shows a plot of the estimated percentage increase from these negative binomial regressions, revealing a stark increase in alcohol-related domestic abuse on days of an England victory, starting in the three hour period of the match, peaking in the three-hour period afterwards, and gradually declining to its original level in the twenty-four hours following the victory.

Figure 5.3: The temporal dynamics of the football-induced increase in domestic abuse, by alcohol involvement



Note: Estimates are from two separate negative binomial regressions (based on tests of overdispersion) with year, month, day of week, three-hour period of day, Christmas, New Year's eve controls. Shaded area is 95% CIs.

These results strongly suggest that the emotional effect of a win drives the subsequent increase in alcohol-related domestic abuse, and highlight the possibility that the effect of England victories stem from prolonged post-match celebrations coupled with increased alcohol consumption. Interestingly, we also see a slight increase in non-alcohol related incidents twelve hours after a loss or a victory, probably reflecting the small increase in non-alcohol related domestic abuse after an England match day seen in Table 5.1.

5.3.2 Reconciling the evidence about the link between football and domestic abuse

Our results have shown that an England victory in a national football tournament is followed by a 61% increase in the reported number of alcohol-related domestic abuse cases. This is a large effect, translating into a 0.43 increase in the daily rate of alcohol-related cases per 100,000 individuals, against a base rate 0.71 cases per 100,000. The effect is entirely limited to alcohol-related abuse, even though alcohol-related domestic abuse cases constitute only 23% of all domestic abuse in our dataset. As such, we see this as strong quantitative evidence for the instrumental role of alcohol in the relationship between football and domestic abuse in England. The effect is also exclusively limited to male-perpetrated domestic abuse, implicating masculinity and alcohol consumption as the pathway by which football increases abuse. The temporal pattern of the increase following an England victory is highly consistent with a causal explanation, further supported by the fact that the allocation of England win days can be largely considered random.

Our findings show both similarities and differences with results from previous quantitative investigations. Replicating the results of a previous US study, we found that it is male to female abuse that is affected by a sporting event (Card & Dahl, 2011). In the same study, the effect of the match did not depend on alcohol-involvement in the abuse case, and the increase was driven by unexpected losses. In contrast, we find that in the context of England and football, it is a victory that results in the largest increase, and that alcohol involvement is critical. This discrepancy most likely stems from the contextual differences between the two studies (England, football, national tournaments vs. US, American football, NFL matches), suggesting that the effect of sports-induced emotional cues on domestic abuse is highly sensitive to the cultural context.

Based on the pre-match betting odds, all England victories were expected in our dataset. This suggests that in the context of England's participation in national football tournaments, it is living up to the expectations of the fans that results in the largest emotional effect. Indeed, English newspapers' narratives about the national team's performance in these tournaments are characterised with high levels of optimism, expectation and yearning for the glory of the 1966 World Cup (Vincent, Kian, Pedersen, Kuntz, & Hill, 2010). Previous research has demonstrated how the vicarious experience of watching their team play can increase supporter's testosterone and cortisol levels, even when they expect their team to win, which has been suggested to be an adaptive response to the perceived threat to one's social identity (van der Meij et al., 2012). Anecdotal evidence suggests that alcohol consumption increases following an England victory (Davies, 2018), consistent with our findings.

The most widely-discussed England-based investigation of the link between football and domestic abuse found that an England loss results in the most pronounced increase in domestic abuse (38%), and a win or draw have a slightly smaller effect (26%; Kirby et al., 2014). This study used daily data on IPV from Lancashire Constabulary (serving a population of 1.4 million, about half the population of the West Midlands) from the two-month period of the 2002, 2004 and 2010 World Cup tournaments (June–July).

Using daily domestic abuse data from the West Midlands for the period between 2010 and 2018, we find a markedly different pattern, with the largest increase in alcohol-involved cases of abuse when England wins, and no evidence for an increase when England loses. Upon re-analysing their data by treating wins and draws as two separate variables (resulting in an improved model fit, see Table 5.5), we see a roughly similar effect for wins (45%, 95% CI [28%–64%]) and losses (39%, 95% CI [18%–64%]), and no effect when England draws. Our reanalysis replicates the win effect seen in our dataset, though the absence of a loss effect remains a stark difference between the two studies. While our sample sizes are different (92 days versus 3,017 days), and our respective samples cover different geographical areas and time periods, the discrepancy is still puzzling.

Table 5.5: Replication of Kirby et al. (2014) with an alternative specification

	Dependent variable:	
	Number of reported IPV cases per day	
	Original Model	Win/Draw Separate
	(1)	(2)
England windraw	0.256***	
	(0.055)	
England win		0.452***
		(0.064)
England draw		0.032
		(0.073)
England loss	0.382***	0.388***
	(0.094)	(0.085)
After England	0.111**	0.113**
	(0.051)	(0.047)
Number of days	92	92
AIC	714.980	704.356

[a] **p<0.1; ***p<0.05; ****p<0.01*

[b] *Estimates are from a series of negative binomial regressions (based on tests of overdispersion) with year and day of week controls; standard errors in parentheses; data are only available during the tournament period*

To explore the underlying reason for this difference and test the robustness of our results, we find it instructive to break our analysis into specific tournament years for the two datasets (see Table 5.6). An interesting common pattern in both samples is the large effect of England's victory over Slovenia in the group stage of the 2010 World Cup, which, after much anticipation, secured their progression to the next stage of the tournament. Equally, the subsequent loss against Germany in the knockout stage resulted in a considerable increase in the number of reported domestic abuse incidents, which is the only tournament in our dataset where this pattern appears. Interestingly, an earlier examination of the 2010 World Cup found a similar pattern, using daily data from 33 out of 39 police forces in England (Brimicombe & Cafe, 2012), although our much larger sample size (3,017 days versus 62 days) allows for a more precise assessment of the link between football and domestic abuse.

Table 5.6: Year subgroup regressions, Lancashire and West Midlands data

	Dependent variable:							
	Number of reported IPV cases per day in Lancashire			Number of reported domestic abuse cases per day in West Midlands				
	negative binomial		*Poisson*				*negative binomial*	
	2002	2006	2010	2010	2012	2014	2016	2018
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Tournament on				0.074^{*}	-0.066	-0.048	0.035	0.089^{*}
				(0.041)	(0.085)	(0.044)	(0.041)	(0.044)
England win	0.596^{***}	0.297^{***}	0.916^{***}	0.050	-0.237		-0.008	0.061
	(0.152)	(0.077)	(0.114)	(0.155)	(0.175)		(0.151)	(0.077)
England draw	0.100	0.098	-0.137	-0.029	0.324	-0.077	-0.021	
	(0.150)	(0.156)	(0.095)	(0.112)	(0.204)	(0.173)	(0.108)	
England loss	0.200	0.373^{***}	0.568^{***}	0.174	-0.127	-0.042	-0.155	0.066
	(0.232)	(0.117)	(0.106)	(0.140)	(0.212)	(0.124)	(0.154)	(0.088)
After England	0.253^{**}	0.122^{*}	0.024	0.070	-0.008	0.007	0.038	0.140^{**}
	(0.101)	(0.070)	(0.065)	(0.082)	(0.125)	(0.103)	(0.081)	(0.060)
Tournament on:Alcohol				-0.093	0.076	0.063	-0.163^{**}	-0.068
				(0.101)	(0.162)	(0.076)	(0.072)	(0.078)
England win:Alcohol				2.558^{***}	0.756^{*}		0.348	0.460^{***}
				(0.277)	(0.314)		(0.257)	(0.123)
England draw:Alcohol				0.078	-0.581	0.089	0.129	
				(0.246)	(0.571)	(0.307)	(0.180)	
England loss:Alcohol				0.748^{**}	0.301	0.048	-0.289	0.160
				(0.259)	(0.372)	(0.206)	(0.322)	(0.149)
After England:Alcohol				0.128	-0.072	0.068	-0.112	0.188^{*}
				(0.183)	(0.254)	(0.171)	(0.144)	(0.102)
Number of days	30	32	30	365	366	365	366	309

[a] *$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$*

[b] *Estimates are from a series of negative binomial or poisson regressions (based on tests of overdispersion). The first three regressions have day of week control, the rest of the regressions have month, day of week, Christmas, New Year's eve controls interacted with alcohol; standard errors in parentheses*

While the effect of a victory or loss is likely to be highly specific to the context of a particular match (e.g., group stage or knockout stage, previous performance of the team, weather on the day, etc.), the estimated effect of an England victory on the number of reported domestic abuse cases is robust to different model specifications (see Table 5.1), using data from a different geographical area (see Table 5.6), and the exclusion of specific tournament years (see Table 5.7).

Table 5.7: Robustness of the result: sensitivity to the exclusion of specific years

	Dependent variable:				
	Number of reported domestic abuse cases per day				
	2018 excluded	2016 excluded	2014 excluded	2012 excluded	2010 excluded
	(1)	(2)	(3)	(4)	(5)
Alcohol	−0.862***	−0.862***	−0.862***	−0.863***	−0.867***
	(0.033)	(0.033)	(0.032)	(0.031)	(0.033)
Tournament on	0.018	0.015	0.027	0.030	−0.003
	(0.022)	(0.025)	(0.025)	(0.022)	(0.025)
England win	−0.093	−0.047	−0.029	0.019	−0.051
	(0.097)	(0.068)	(0.062)	(0.066)	(0.067)
England draw	0.038	0.077	0.057	0.004	0.046
	(0.072)	(0.091)	(0.078)	(0.075)	(0.088)
England loss	0.030	0.066	0.053	0.054	0.013
	(0.079)	(0.065)	(0.069)	(0.062)	(0.065)
After England	0.057	0.080*	0.088**	0.099**	0.071*
	(0.048)	(0.042)	(0.040)	(0.039)	(0.042)
Alcohol:Tournament on	−0.086**	−0.037	−0.118***	−0.092**	−0.048
	(0.039)	(0.046)	(0.047)	(0.040)	(0.042)
Alcohol:England win	0.884***	0.674***	0.609***	0.574***	0.511***
	(0.163)	(0.109)	(0.100)	(0.105)	(0.107)
Alcohol:England draw	−0.046	−0.141	−0.048	0.055	−0.017
	(0.130)	(0.179)	(0.141)	(0.131)	(0.151)
Alcohol:England loss	0.014	0.139	0.131	0.078	0.039
	(0.134)	(0.107)	(0.116)	(0.103)	(0.109)
Alcohol:After England	−0.065	0.096	0.050	0.054	0.050
	(0.086)	(0.073)	(0.071)	(0.067)	(0.071)
Number of days	2,708	2,651	2,652	2,651	2,652

[a] **p*<0.1; ***p*<0.05; ****p*<0.01*

[b] *Estimates are from a series of negative binomial regressions (based on tests of overdispersion) with year, month, day of week, Christmas, New Year's eve controls interacted by alcohol; standard errors in parentheses*

5.3.3 Extensions: rugby, other abusive behaviours

Does this effect generalise to other sporting events, or is it specific to football? It has been previously suggested that other popular sports, such as rugby have similar links with domestic abuse (Brooks-Hay & Lombard, 2018). Rugby is the second most popular sport in England after football (Ipsos MORI, 2003). Focusing on the Six Nations, a high-profile rugby tournament that takes place every year with the participation of England, Wales, Scotland, Ireland, France and Italy, we explored whether the reported number of domestic abuse cases increase on days when the England national rugby team plays. Between 2010 and 2018, there are many more win and loss days of the England rugby union team compared to the England national football team, providing us with more statistical power to identify a potential effect. The results show no comparable effects for rugby matches (see Table 5.8), possibly stemming from differences in media coverage, audience numbers, and the role of alcohol between the two tournaments.

We also investigated whether England match days have similar effects on other types of abusive behaviours, including sexual offences, child and vulnerable adult abuse. A commonality between domestic abuse and these types of offences is the element of control and domination, although domestic abuse covers a much wider range of behaviours and is consequently significantly more frequent in our dataset. We find no evidence that England matches have comparable effects on non-domestic sexual offences and other abuse cases (see Table 5.9).

Table 5.8: The effect of England matches in the Six Nations rugby tournament on domestic abuse

	Dependent variable:
	Number of reported domestic abuse cases per day
Alcohol	-0.862^{***}
	(0.031)
Tournament on	0.005
	(0.019)
England win	0.0001
	(0.035)
England loss	0.056
	(0.055)
After England	-0.010
	(0.031)
Alcohol:Tournament on	-0.047
	(0.035)
Alcohol:England win	0.045
	(0.059)
Alcohol:England loss	-0.073
	(0.091)
Alcohol:After England	-0.021
	(0.055)
Number of days	3,017

[a] $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

[b] *Estimates are from a series of negative binomial regressions (based on tests of overdispersion) with year, month, day of week, Christmas, New Year's eve controls interacted by alcohol; there was only one England rugby match that resulted in a draw between 2010 and 2018, therefore we excluded it from the data; standard errors in parentheses*

Table 5.9: Non domestic abuse incidents that are about power

	Dependent variable:	
	Number of reported cases per day	
	Sexual Offences	Other Abuse
	(1)	(2)
Alcohol	-0.955^{***}	-0.950^{***}
	(0.146)	(0.075)
Tournament on	0.079	0.078^{*}
	(0.068)	(0.042)
England win	-0.172	-0.073
	(0.217)	(0.132)
England draw	-0.062	0.175
	(0.253)	(0.148)
England loss	-0.220	0.153
	(0.223)	(0.132)
After England	-0.035	0.095
	(0.134)	(0.081)
Alcohol:Tournament on	-0.121	-0.069
	(0.157)	(0.093)
Alcohol:England win	0.191	0.166
	(0.462)	(0.274)
Alcohol:England draw	0.781	-0.252
	(0.503)	(0.346)
Alcohol:England loss	0.011	-0.111
	(0.483)	(0.285)
Alcohol:After England	0.114	-0.172
	(0.287)	(0.182)
Number of days	3,017	3,017

[a] *$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$*

[b] *Estimates are from a series of negative binomial regressions (based on tests of overdispersion) with year, month, day of week, Christmas, New Year's eve controls interacted by alcohol; standard errors in parentheses*

5.3.4 Characteristics of domestic abuse perpetrated on England match days

Our data further allows us to explore the characteristics of alcohol-related domestic abuse perpetrated on England match days. First, using a series of logistic regressions, we investigate whether these cases are more likely to be newly reported (with no earlier record for the same victim-offender pair in our dataset), happen in a residential dwelling as opposed to a public location, or result in an injury. We find no evidence that domestic abuse cases perpetrated on England match days are more likely to be newly reported (see Table 5.10), compared to domestic abuse cases occurring on non-match days. It could be argued that since fans often congregate in pubs to watch England play, there is a higher likelihood that domestic abuse occurs in public and get reported on these days. Interestingly, our results indicate that, compared to non-match days, reported cases are more likely to be perpetrated in public on England loss days, but not on England win days, and that this effect does not differ by alcohol-involvement in the case. Non-alcohol related cases reported on England loss days are also more likely to result in an injury, a pattern that is absent from alcohol-related cases.

Next, we turn to repeated cases of domestic abuse (multiple cases with the same victim-offender pair). Domestic abuse is rarely a one-off incident, and reported repeat cases allow us to explore the characteristics of domestic abuse that occurs on match days in more detail. We are interested in whether the number of days elapsed between two consecutive cases is affected by England football matches. For example, it is possible that England match days bring reported cases of domestic abuse forward, which would have otherwise happened at a later point in time. We investigate this question with two negative binomial regressions, where the outcome variables are the number of days elapsed since the last reported case, and the number of days until the next case, respectively.

Table 5.10: Characteristics of domestic abuse cases reported on match days I

	Dependent variable:		
	Newly Reported Yes=1, No=0	Public Location Yes=1, No=0	Results in Injury Yes=1, No=0
	(1)	(2)	(3)
Alcohol=Yes	−0.030	0.001	0.427***
	(0.059)	(0.081)	(0.058)
Tournament on	−0.037	0.021	0.007
	(0.030)	(0.037)	(0.033)
England win	0.011	0.167	0.153
	(0.089)	(0.110)	(0.101)
England draw	0.082	0.014	0.119
	(0.121)	(0.138)	(0.117)
England loss	−0.099	0.337***	0.265***
	(0.086)	(0.099)	(0.093)
After England	0.035	0.070	0.049
	(0.056)	(0.068)	(0.062)
Alcohol:Tournament on	0.087	0.063	−0.058
	(0.060)	(0.080)	(0.066)
Alcohol:England win	0.093	0.104	−0.064
	(0.156)	(0.196)	(0.165)
Alcohol:England draw	−0.151	−0.016	−0.209
	(0.233)	(0.306)	(0.237)
Alcohol:England loss	0.221	0.044	−0.413**
	(0.171)	(0.198)	(0.182)
Alcohol:After England	−0.036	0.042	−0.122
	(0.108)	(0.143)	(0.118)
Number of cases	251,976	279,777	279,777

[a] *$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$*

[b] *Estimates are log odds from a series of logistic regressions with year, month, day of week, Christmas, New Year's eve controls interacted by alcohol, where every observation is a reported domestic abuse case; cases that happened in 2010 were excluded from the first regression; standard errors clustered by victim-offender pairs are in parentheses*

Table 5.11: Characteristics of domestic abuse cases reported on match days II

	Dependent variable:		
	Days since last	Days until next	Hours until reported
	(1)	(2)	(3)
Alcohol	0.103^{*}	0.057	-0.839^{***}
	(0.054)	(0.050)	(0.189)
Tournament on	-0.014	-0.047^{*}	0.080
	(0.028)	(0.028)	(0.063)
England win	0.016	-0.340^{***}	-0.098
	(0.082)	(0.095)	(0.162)
England draw	-0.017	-0.111	0.034
	(0.096)	(0.105)	(0.208)
England loss	-0.163^{*}	-0.104	-0.560^{***}
	(0.087)	(0.087)	(0.170)
After England	0.052	-0.139^{**}	-0.243^{**}
	(0.054)	(0.055)	(0.108)
Alcohol:Tournament on	0.026	0.025	0.200
	(0.057)	(0.056)	(0.197)
Alcohol:England win	-0.119	0.358^{**}	0.152
	(0.146)	(0.159)	(0.450)
Alcohol:England draw	-0.266	-0.116	-0.935^{**}
	(0.231)	(0.208)	(0.390)
Alcohol:England loss	0.277^{*}	0.114	0.552
	(0.159)	(0.166)	(0.654)
Alcohol:After England	-0.104	0.147	-0.265
	(0.106)	(0.102)	(0.297)
Number of cases	95,091	95,091	272,793

[a] *$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$*

[b] *Estimates are from a series of negative binomial regressions (based on tests of overdispersion) with year, month, day of week, Christmas, New Year's eve controls interacted by alcohol, where every observation is a reported domestic abuse case; for each regression, we excluded the upper 2.5% of the outcome variable; standard errors clustered by victim-offender pairs are in parentheses*

In addition, using all reported cases, we explore whether the number of hours elapsed before reporting the case is affected by England match days.

The results show that non-alcohol related cases perpetrated on England loss days occur fewer days after the previous incident, 192 days, 95% CI [159–232 days], compared to non-alcohol repeat cases reoccurring on non-match days, 226 days, 95% CI [207–248 days] (see Table 5.11). Non-alcohol related domestic abuse cases perpetrated on England win days are more likely to be followed by another case of abuse in fewer days, 172 days, 95% CI [138–214 days], compared to cases occurring on non-match days, 242 days, 95% CI [223–261 days], and this pattern is absent from alcohol-related cases. Interestingly, non-alcohol related cases perpetrated on England loss days are likely to be reported after fewer hours, 59 hours, 95% CI [45–78 hours], compared to non-alcohol related abuse perpetrated on non-match days, 104 hours, 95% CI [91–119 hours].

Finally, using the sample of repeated cases, we explore whether previously non-alcohol related cases are more likely to reoccur as alcohol-related abuse on England match days. We investigate this question using a logistic regression, whilst controlling for the type of the previous case (alcohol/non-alcohol related). The results show that on England win days, there is an increased likelihood of an alcohol-related case occurring, irrespective of whether the previous case was alcohol-related or not (see Table 5.12). Taken together, these results indicate that apart from the higher likelihood of alcohol-involvement, domestic abuse that follows an England victory is not characteristically different from domestic abuse perpetrated on other days during the year.

Table 5.12: Alcohol transition on England match days

	Dependent variable:
	Alcohol-involvement in case Yes=1, No=0
Tournament on	-0.134^{**}
	(0.062)
England win	0.443^{***}
	(0.157)
England draw	0.368^{*}
	(0.201)
England loss	-0.113
	(0.180)
After England	0.041
	(0.114)
Tournament on:Previous alcohol	-0.051
	(0.100)
England win:Previous alcohol	-0.110
	(0.277)
England draw:Previous alcohol	-0.365
	(0.372)
England lost:Previous alcohol	0.179
	(0.292)
After England:Previous alcohol	0.066
	(0.180)
Number of cases	97,292

[a] *$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$*

[b] *Estimates are log odds from a logistic regression with year, month, day of week, Christmas, New Year's eve controls interacted by alcohol involvement of the previous case, where every observation is a reported domestic abuse case; standard errors clustered by victim-offender pairs are in parentheses*

5.4 General Discussion

Previously, a number of quantitative and qualitative studies have explored the link between England's participation in national football tournaments and domestic abuse. To our knowledge, our analysis is the most extensive quantitative investigation of this relationship to date. Our data, consisting of 9 years of crimes and incidents recorded by the third largest police force in England, allowed us to analyse various aspects of the relationship between football and domestic abuse.

To summarise, we have found that when the England national football team wins, there is a 61% increase in alcohol-related domestic abuse, primarily driven by male-to-female abuse. An increase is also seen in other violent crimes, predominantly in violence perpetrated by men against women. The temporal pattern of the increase suggests a causal mechanism, and the effect is robust to the exclusion of specific tournament years, and using data from a different time period and geographical area within England. The effect is specific to football, but not rugby. Apart from the higher likelihood of alcohol-involvement, abuse cases occurring on England win days are not characteristically different from abuse occurring on other days throughout the year.

In contrast with this, while we found no evidence for an increase in the reported number of domestic abuse cases on England loss days (see Table 5.1), domestic abuse perpetrated on these days seems to be characteristically different from domestic abuse perpetrated on other days in that these cases are more likely to occur outside, result in an injury, and get reported sooner. Furthermore, repeated cases perpetrated on England loss days occur slightly sooner following the previous case, but abuse perpetrated on England win days is followed by another incident sooner. While these findings should be interpreted with caution due to the pervasive problem of underreporting, the results suggest important differences in the effect of England wins and losses on domestic abuse. In particular, while we found no increase in the overall number of cases reported on England loss days, incidents reported on these days are characteristically different from abuse perpetrated on non-match days. In contrast, we observe a substantial difference in the number, but not in the characteristics of cases perpetrated on England win days.

It has been previously suggested that the link between football and domestic abuse is not causal, and that other, alternative explanations can account for this apparent relationship, including increased policing on England match days, the effect of awareness campaigns before the tournaments, and other high-profile events taking place around the time of the match (Brooks-Hay & Lombard, 2018). We could expect that higher levels of policing on England match days would result in an increase in the number of recorded cases perpetrated outside, and that a successful pre-tournament awareness campaign would result in an increase in the number of newly reported cases. Our results do not support either of these alternative hypotheses (see Table 5.10). In addition, it is unclear why the effect of other events, different policing practices, or awareness campaigns would depend on the result of the match.

In contrast, we argue that our results provide strong evidence for a causal link between football and domestic abuse for several reasons. First, the allocation of England match days are random, as they are determined through a draw. In addition, the outcome of any individual match is getting increasingly less predictable as the tournament progresses on. But more importantly, our three-hour analysis of the England win effect (Figure 5.3) show that the temporal pattern of the effect is highly consistent with a match-induced explanation of the increase, making it unlikely that other events occurring on England win days would be responsible for the increase.

Our results show that alcohol plays an instrumental role in the nexus between football and domestic abuse in England. While we are not able to quantify how alcohol consumption changes during national football tournaments, anecdotal evidence suggests a substantial increase (Gornall, 2014; Davies, 2018; Fraser

McKevitt, 2018). We speculate that the extent of this increase is likely to depend on the range of factors, most notably England's performance in the tournaments, and the weather on England match days. The importance of alcohol in the context of England's World Cup participation and its association with reported domestic abuse cases is in clear contrast with US-based evidence regarding the link between NFL games and IPV (Card & Dahl, 2011), which suggested that the match-induced increase in the number of reported IPV cases does not depend on alcohol-involvement in the case.

This difference is likely to be a manifestation of cultural differences in the popularity and nature of sports-related alcohol consumption in the two countries. In the UK, drinking culture is embedded in football, through the practices of sponsorship and targeted advertising (Alcohol Concern, 2014). It is estimated that televised top-class English professional football matches feature a visual or verbal alcohol reference every two minutes (Graham & Adams, 2013), highlighting the strong connection between alcohol and the football spectator experience. Illustrating the strength of this link, the Home Office banned fans from consuming alcohol in view of the pitch in 1985 (Alcohol Concern, 2014).

While our results show a clear relationship between alcohol consumption, football and domestic abuse in England, it is important to stress that neither televised football nor alcohol consumption causes domestic abuse. In doing so, we wish to clarify that we are not using football and alcohol as a "scapegoat" for domestic abuse (e.g., Riordan, 2018; Brooks-Hay & Lombard, 2018). Importantly, we do not suggest that football and alcohol turns previously non-abusive people into abusers (in fact, our results show that this is clearly not the case), or that domestic abuse only happens on match days.

However, we do believe that the patterns observed in our dataset are strongly consistent with a victory-induced increase in the number of reported domestic abuse cases, and are not due to other factors as suggested by Brooks-Hay and Lombard (2018). In our view, the most likely explanation for this observed increase is that England football matches, and in particular, victory celebrations serve as an opportunity for abusive individuals to be violent. Understanding the exact pathways of this phenomenon remains an important question for future research.

For victims, domestic abuse does not occur once every four years following a football match, but is a lived experience of constant fear (Brooks-Hay & Lombard, 2018). Nevertheless, our results provide a deeper understanding of the contexts that can be conducive to abuse. In particular, these findings illuminate that the experience of "national success" in a highly male-dominated sport is a breeding ground for male-perpetrated, alcohol-related domestic abuse.

From a policy perspective, it is hard to efficiently tackle such a complex phenomenon. In addition to the UK Government's substantial ongoing efforts to fight domestic abuse (Home Office, 2016), leaders in football can do much more to improve the reputation of British football fandom, and make it more inclusive for women and minorities. For example, consciously aiming to make football more attractive to women and increase their presence in this historically male-dominated arena can be an effective way to fight sexist, misogynistic attitudes that are too often evident amongst male football fans (Woodward, 2017), and also underpin the world view of many domestic abuse offenders (Peralta & Tuttle, 2013). At the same time, football clubs need to be explicit and strict in condemning violent, abusive behaviours of their fans, and take part in raising awareness of the dark side of the "beautiful game" (Swallow, 2017). Such measures can be important first steps towards a radical transformation of football culture and form the basis of a meaningful change.

Chapter 6

Conclusion

This thesis presented four research projects in the topic of context effects influencing human behaviour, and showcased examples from the wide range of interdisciplinary methodological approaches used in behavioural economics. In Chapters 2–4, we focused on violations of the rationality assumption in value-based choice, using laboratory-based experimental methods. More specifically, Chapter 2 contrasted various forms of context-dependent choice behaviour within a popular process model of choice, while Chapters 3 and 4 focused on exploring the boundary conditions of a well-known cognitive bias in choice, the attraction effect. In contrast to these laboratory-based investigations of human behaviour, Chapter 5 presented a statistical analysis of a large observational dataset, and explored the link between football and domestic abuse in the West Midlands. Below we summarize the findings, limitations, and possible future directions that naturally follow from each of these investigations.

In Chapter 2, our aim was to compare four distinct forms of subjective valuation (range, rank, local, and global value transformation) underlying choice behaviour. We chose to explore these four subjective value transformation rules within the modelling framework of the DDM, and its extension, the aDDM, using data from two trinary choice experiments, Experiment 1, which was a value-based choice experiment (where we also recorded participants' eye movements during choice), and Experiment 2, which was a perceptual version of Experiment 1. We conducted three empirical tests to compare the explanatory power of these four subjective transformation rules. First, using an improved, simulations-based model fitting approach and data from Experiment 1, we identified the subjective transformation rule that provided the best fit for each participant's choice data. Second, we repeated the same test, but instead of a simulations-based approach, we used a novel probability distribution method to derive choice probabilities. The results from these two comparisons were remarkably similar, and indicated that for the majority of participants (about 57%), subjective valuation was guided by the absolute value of the options, while for a smaller proportion of participants (about 20%), choice behaviour could be best described by the relative valuation rule.

In a third comparison, we investigated the explanatory power of the four rules by comparing their ability to predict the qualitative changes in choice proportions in response to a set of pre-determined choice set manipulations. For this comparison, we used data from both a value-based (Experiment 1) and a perceptual choice (Experiment 2) experiment, to further explore the domain-specificity of context-dependent choice behaviour. This comparison has revealed that only the global and local maximum rules could correctly predict the effect of three out of the four choice set manipulations, and thus strongly supported the results of the two previous tests.

Therefore, our results have shown that choice behaviour in these experiments can be best described by a hybrid valuation mechanism, where subjective valuation is mostly affected by the absolute value of the items, and to a lesser extent, also by their relative values. These findings were remarkably similar across two very different stimuli domains. In addition, our novel approach to derive choice probabilities delivered very similar results to those obtained from a simulations-based method, alleviating concerns about employing a simulations-based estimation approach to derive choice probabilities in a stochastic model framework.

These results contribute to the growing body of research investigating economic decision making within a process model framework. A clear direction for future research could involve estimating appropriate weights for a hybrid subjective valuation model within the DDM framework. However, a limitation of our study is that we only focused on one broad type of sequential sampling model (the DDM, and its extension the aDDM), and one specific way to capture the idea of context dependency within this modelling framework (through the items' values in the evidence accumulation equations). Future investigations could probe the sensitivity of the results, by exploring alternative ways to incorporate the idea of context sensitivity within the DDM (e.g., through modelling changes in the drift rate, or changing the relative evidence accumulation equations) or in the framework of other sequential sampling models. Through directly measuring neural firing rates during choice, the rapidly expanding field of neuroeconomics will undoubtedly deliver important insights about biologically plausible ways to incorporate context dependence within various sequential sampling models, including the DDM.

In Chapter 3, our aim was to explore the boundary conditions of the attraction effect, by testing whether this well-known decision bias extends to choice scenarios with complex, naturalistic choice options. In Experiment 1, we used movie posters and their associated summary texts as inputs to a LSA algorithm to uncover the topics underlying our selected set of movies, and establish the similarity of each movie pair. Drawing on insights from a previous debate about the methodological challenges associated with testing the attraction effect with real-world stimuli, we designed the first truly rigorous test of the attraction effect with naturalistic choice options. In Experiment 1, we found no evidence for the attraction effect, but our data has also revealed that participants did not perceive our target-decoy pairs as sufficiently similar. In Experiment 2, we circumvented this problem with a more careful selection of target-decoy pairs, but still found no evidence for the attraction effect.

One limitation of this investigation is that we only used one type of real-world stimuli to probe the attraction effect. However, given the rigour of our methodological approach, we see these results as sound evidence supporting the claim that the attraction effect does not extend to choices with naturalistic options. Historically, the attraction effect played a key role in the development of a number of very influential choice models, which claim to be able to capture context dependence in decision making. Our results cast serious doubt on the appropriacy of relying on the attraction effect in the development of choice models aiming to offer a realistic account of real-world decision making.

In both of these chapters, we used movies posters as choice items: they served as an example of a preferential choice task in Chapter 1 (as opposed to a perceptual choice task), and as an example of a naturalistic, real-world choice item in Chapter 2 (as opposed to a choice item with numerical attribute dimensions). Whether our findings generalise to other types of stimuli is an important question. Our movie stimulus is visually rich, and thus superficially similar to a lot of the everyday consumer products people choose between on a daily basis. In addition, we obtained similar results from a preferential and perceptual choice experiment in Chapter 2, indicating that context effects might be insensitive to the exact nature of stimuli.

However, movie posters differ from the most common choice items people encounter in at least two respects. First, in our experiments, we did not incorporate any price information. As movies are often priced the same, this information is unlikely to be of crucial importance for this particular choice item, however, price information is a fundamental attribute of everyday consumer items. Importantly, if context effects are more likely to arise in choices where there are numerical attributes (owing to the existence of one highly salient comparison dimension), then this might mean that our results can be considered a strict test of context sensitivity, and could also explain why we have not found strong evidence for context-dependent behaviour in these experiments.

In addition, our movie stimulus is highly symbolic, which is not the case for most consumer choice items, where the specific attributes of the product are easily accessible, thus reducing the complexity of the preference construction process. Movie posters are visually rich, and complex mental processes are likely to play an important part in their preferential evaluation and subsequent comparison. Given that context effects have mostly been demonstrated in choice scenarios involving non-symbolic, simple stimuli, it is entirely possible that the complex nature of our chosen stimulus introduced confounding factors we could not control for, subsequently masking any context effects that would otherwise be more apparent in choices involving simpler choice items.

Understanding how preference formation depends on the attribute properties and symbolic nature of the stimuli is an interesting avenue for future research. It would possible to test whether preferences are dependent on the exact representation of a complex stimuli (e.g., movies can be represented by just their title, or by scores along genre dimensions), and explore how this might affect susceptibility to context effects in preferential choice.

Our results from Chapter 3 posed an interesting question about whether the attraction effect is simply the product of the cognitive comparison strategy used in choices involving numerical attributes. We therefore decided to further investigate this question in Chapter 4. Based on findings regarding the presence of the attraction effect in perceptual choice experiments, and our previous findings, we conjectured that a sequential, attribute-wise comparison process is key in producing the attraction effect. The purpose of Chapter 4 was to test this hypothesis. Drawing on decades of psychological research on information processing, we designed a within-subjects experiment with two versions of the same task, one with spatially separable numerical attributes (the standard task used in the literature for testing the attraction effect), and a pictorial version, where the two attributes were integral, and thus spatially non-separable. To replicate the conditions under a value-based choice situation with our artificially created stimuli, participants were required to learn a valuation rule in a learning stage that preceded the choice stages, and were further instructed to rely on this rule in the subsequent choice trials. Our results were mixed. First, we found no evidence for the attraction effect in the pictorial condition, which was predicted. Second, we found a strong effect of the order of the two types of tasks, such that we only saw a strong attraction effect in the numerical condition when it followed the pictorial version, which was in stark contrast with our expectations.

The limitations of our experiment are twofold. First, it could be argued that our approach to include a learning stage to induce preferences over our abstract stimuli set to study value-based choice did not manage to invoke the same cognitive processes that underlie a natural, simplistic preferential choice scenario. Other approaches involving stimuli that have naturally interpretable attribute dimensions are more promising in understanding what gives rise to the attraction effect. Second, the learning stages in our experiment turned out to be harder and more cognitively demanding than we anticipated, further complicating the interpretation of the results. Unfortunately, we did not collect enough data to understand the exact cognitive mechanisms driving our results, but our findings highlight the sensitivity of the attraction effect to the stimulus domain and task characteristics. A future research project could conduct a more comprehensive investigation of the link between stimulus presentation and the strength of the attraction effect, by testing how different degrees of attribute separability modulate the attraction effect.

In contrast with the laboratory-based, experimental approach used in previous chapters, Chapter 5 used a real-world, observational dataset to investigate the relationship between England's participation in national football tournaments and the number of reported domestic abuse incidents. Our results have demonstrated a substantial, 61% increase in the number of reported alcohol-related domestic abuse incidents following an England victory. While it is extremely hard to establish causality in the context of an observational dataset, a three-hour analysis showed that the timeline of the effect is highly consistent with a causal link between football and domestic abuse. Our results are in line with previous, seemingly contradictory findings. Our main contribution lies in the fact that in the context of England, our study is the first to explore the instrumental role of alcohol in the nexus between football and domestic abuse.

Our results have demonstrated an important example of how visceral factors and alcohol consumption can create an environment that is conducive to domestic abuse. A limitation of our study is that we have only used data from one county of England from a specific time period, and thus future replication attempts involving different regions and time periods will be important in probing the robustness of this effect. We also recognise that domestic abuse is a complex, multifaceted phenomenon, and that football and alcohol alone do not cause domestic abuse. However, we believe that qualitative studies exploring the complex link between football fandom, alcohol consumption and attitudes towards women amongst domestic abuse perpetrators will be key in elucidating the pathways of this phenomenon, and understanding its specific cultural drivers. In addition, we believe efforts focused on increasing women's participation and visibility in British football fandom can be an important first step to dissociate football from male-perpetrated domestic abuse.

Overall, the research projects described in this thesis demonstrated examples from the wide range of contextual factors affecting human behaviour, and illustrated some of the eclectic methodological approaches used by behavioural economics to investigate these factors. In doing so, it hoped to at least partly reveal what makes behavioural economics such an immensely successful and rapidly growing academic field.

The reunification of economics and psychology in the form of behavioural economics was necessary, if not inevitable. Human behaviour is shaped by a myriad of external and internal factors, and often seems inconsistent and unpredictable. This makes the development of reliable predictive models of human decision making extremely challenging. For a long time, economics attempted to explain seemingly contradictory behavioural phenomena by constantly amending its theoretical models, whilst ignoring decades of knowledge accumulated from psychological research on exactly the same phenomena, on the basis of their theoretical and methodological differences. The emergence of behavioural economics has changed this disciplinary status quo, and initiated a slow revolution within economics by moving interdisciplinary initiatives from the periphery to the mainstream.

The key to the success of behavioural economics lies in its inherently interdisciplinary nature. Crossing traditional disciplinary boundaries allowed researchers to amalgamate the theoretical and methodological strengths of multiple disciplines, creating a novel field with the ability to draw on a much wider range of scientific knowledge. Inevitably, this fusion lead to the rapid accumulation of new insights about choice phenomena that had been studied for decades. Technological developments have provided an important source of novel methodological tools for behavioural economics, and will continue to do so. In turn, behavioural economics will no doubt keep delivering invaluable insights about the factors shaping human behaviour, owing to its cross-disciplinary openness and methodological flexibility.

Appendices

Appendix A

Table A.1: Best fitting parameter value for each participant and value transformation rule, simulations method

		Global Max			Local Max			Range			Rank		
Participant	Best fitting rule	θ	σ	d	θ	σ	d	θ	σ	d	θ	σ	d
1	Global Max	0.75	0.19	0.32	0.76	0.20	0.26	0.86	0.20	0.10	0.64	0.21	0.12
2	Rank	0.57	0.24	0.71	0.68	0.23	0.32	0.61	0.22	0.18	0.58	0.23	0.20
3	Local Max	0.68	0.22	0.49	0.68	0.22	0.36	0.54	0.23	0.22	0.59	0.23	0.21
4	Global Max	0.66	0.26	0.30	0.66	0.26	0.21	0.59	0.26	0.11	0.65	0.26	0.10
5	Global Max	0.76	0.20	0.39	0.75	0.20	0.33	0.86	0.20	0.17	0.84	0.20	0.18
6	Local Max	0.63	0.24	0.32	0.48	0.23	0.31	0.65	0.24	0.21	0.60	0.24	0.22
7	Rank	0.63	0.17	0.22	0.65	0.17	0.20	0.47	0.16	0.15	0.48	0.16	0.15
8	Range	0.78	0.22	0.53	0.77	0.23	0.41	0.61	0.20	0.26	0.55	0.21	0.27
9	Global Max	0.86	0.18	0.21	0.88	0.19	0.17	0.54	0.18	0.15	0.76	0.19	0.12
10	Global Max	0.54	0.23	0.25	0.54	0.23	0.22	0.19	0.23	0.18	0.19	0.24	0.17
11	Global Max	0.67	0.13	0.26	0.64	0.14	0.18	0.64	0.15	0.10	0.59	0.15	0.11
12	Global Max	0.82	0.21	0.31	0.74	0.21	0.27	0.69	0.23	0.15	0.63	0.23	0.16
13	Global Max	0.45	0.19	0.31	0.40	0.19	0.26	0.17	0.20	0.18	0.21	0.20	0.17
14	Global Max	0.42	0.38	0.62	0.25	0.39	0.43	0.17	0.40	0.34	0.23	0.40	0.35
15	Global Max	0.63	0.14	0.18	0.53	0.15	0.16	0.52	0.14	0.11	0.48	0.15	0.11
16	Global Max	0.84	0.18	0.25	0.90	0.18	0.20	0.91	0.18	0.10	0.99	0.19	0.09
17	Global Max	0.79	0.26	0.71	0.79	0.27	0.57	0.59	0.27	0.29	0.48	0.26	0.31
18	Global Max	0.68	0.29	0.58	0.58	0.30	0.42	0.50	0.31	0.27	0.48	0.30	0.28
19	Global Max	0.62	0.19	0.32	0.58	0.18	0.26	0.57	0.18	0.19	0.51	0.18	0.20
20	Global Max	1.05	0.21	0.35	1.00	0.21	0.28	0.96	0.22	0.18	0.95	0.22	0.18
21	Global Max	0.59	0.22	0.32	0.53	0.22	0.31	0.28	0.21	0.28	0.34	0.21	0.26
22	Range	0.88	0.11	0.15	0.84	0.11	0.13	0.59	0.11	0.08	0.49	0.10	0.09
23	Global Max	0.70	0.14	0.15	0.75	0.14	0.12	0.53	0.14	0.09	0.49	0.14	0.10
24	Local Max	1.12	0.25	0.36	1.09	0.24	0.26	1.14	0.25	0.12	1.10	0.25	0.13
25	Local Max	0.22	0.21	0.12	0.20	0.21	0.09	0.22	0.21	0.07	0.24	0.21	0.08
26	Global Max	0.92	0.18	0.32	0.92	0.18	0.26	0.84	0.17	0.18	0.90	0.17	0.17
27	Global Max	0.48	0.12	0.18	0.48	0.13	0.15	0.23	0.14	0.10	0.20	0.14	0.10
28	Local Max	0.18	0.16	0.12	0.19	0.16	0.10	0.18	0.16	0.08	0.18	0.17	0.08
29	Global Max	0.27	0.19	0.28	0.25	0.20	0.18	0.19	0.20	0.16	0.20	0.20	0.16
30	Global Max	0.57	0.29	0.79	0.54	0.29	0.62	0.15	0.31	0.38	0.11	0.31	0.39
31	Range	0.25	0.22	0.01	0.30	0.22	0.00	0.26	0.21	-0.03	0.27	0.22	-0.03
32	Range	0.72	0.20	0.34	0.59	0.19	0.25	0.24	0.20	0.16	0.23	0.20	0.16
33	Local Max	0.78	0.21	0.25	0.68	0.21	0.22	0.57	0.21	0.12	0.57	0.20	0.14
34	Local Max	0.36	0.14	0.18	0.20	0.12	0.14	0.16	0.13	0.10	0.16	0.13	0.10
35	Global Max	0.36	0.32	0.41	0.49	0.32	0.30	0.24	0.33	0.24	0.28	0.32	0.26
36	Range	0.84	0.16	0.22	0.80	0.16	0.22	0.91	0.16	0.10	0.86	0.16	0.11
37	Global Max	0.73	0.16	0.31	0.79	0.16	0.22	0.80	0.17	0.11	0.89	0.17	0.10
38	Local Max	0.57	0.22	0.23	0.56	0.22	0.23	0.23	0.22	0.17	0.24	0.22	0.16
39	Global Max	0.15	0.38	0.23	0.23	0.38	0.11	0.28	0.37	-0.01	0.33	0.37	0.02
40	Global Max	0.70	0.16	0.35	0.64	0.16	0.23	0.26	0.17	0.15	0.24	0.17	0.15
41	Rank	0.97	0.17	0.21	0.97	0.17	0.19	1.03	0.17	0.10	1.09	0.16	0.10
42	Global Max	0.81	0.19	0.33	0.85	0.19	0.25	1.04	0.20	0.12	1.02	0.20	0.11
43	Local Max	0.69	0.19	0.31	0.74	0.19	0.23	0.51	0.21	0.14	0.51	0.21	0.15
44	Global Max	0.60	0.22	0.32	0.52	0.22	0.28	0.42	0.24	0.19	0.44	0.23	0.18
45	Rank	0.97	0.20	0.49	0.96	0.20	0.42	0.76	0.19	0.27	0.81	0.18	0.27
46	Global Max	0.74	0.17	0.30	0.72	0.17	0.23	0.52	0.17	0.15	0.52	0.17	0.15
47	Range	1.02	0.24	0.30	0.99	0.26	0.25	1.06	0.24	0.15	1.02	0.24	0.16
48	Global Max	0.97	0.22	0.34	1.03	0.23	0.29	1.07	0.23	0.12	1.11	0.24	0.11
49	Local Max	0.40	0.20	0.48	0.46	0.19	0.31	0.21	0.21	0.22	0.24	0.21	0.24
50	Local Max	0.82	0.18	0.31	0.76	0.18	0.27	0.68	0.18	0.18	0.86	0.17	0.17

Table A.2: Best fitting parameter value for each participant and value transformation rule, probability distributions method

		Global Max			Local Max			Range			Rank		
Participant	Best fitting rule	θ	σ	d	θ	σ	d	θ	σ	d	θ	σ	d
1	Global Max	0.78	0.19	0.31	0.78	0.19	0.26	0.57	0.20	0.12	0.53	0.20	0.12
2	Rank	0.64	0.24	0.63	0.76	0.22	0.29	0.60	0.22	0.18	0.61	0.22	0.19
3	Local Max	0.70	0.22	0.46	0.68	0.22	0.34	0.50	0.22	0.22	0.50	0.22	0.21
4	Global Max	0.75	0.25	0.26	0.71	0.25	0.19	0.59	0.26	0.10	0.53	0.26	0.10
5	Global Max	0.73	0.19	0.38	0.72	0.19	0.32	0.64	0.19	0.20	0.63	0.19	0.21
6	Local Max	0.67	0.23	0.30	0.59	0.23	0.27	0.54	0.24	0.23	0.49	0.23	0.23
7	Rank	0.67	0.17	0.21	0.65	0.17	0.19	0.53	0.15	0.15	0.54	0.15	0.14
8	Range	0.74	0.21	0.51	0.74	0.21	0.41	0.59	0.19	0.27	0.72	0.21	0.21
9	Global Max	0.83	0.18	0.21	0.79	0.18	0.18	0.70	0.18	0.13	0.68	0.19	0.13
10	Global Max	0.61	0.22	0.23	0.60	0.23	0.20	0.47	0.23	0.13	0.44	0.23	0.12
11	Global Max	0.59	0.13	0.28	0.64	0.13	0.18	0.44	0.14	0.13	0.41	0.15	0.13
12	Global Max	0.74	0.20	0.32	0.74	0.20	0.26	0.63	0.22	0.16	0.59	0.23	0.16
13	Global Max	0.44	0.19	0.28	0.42	0.19	0.24	0.08	0.19	0.19	0.07	0.20	0.19
14	Global Max	0.51	0.36	0.50	0.42	0.37	0.36	0.19	0.37	0.28	0.17	0.37	0.28
15	Global Max	0.56	0.13	0.18	0.54	0.14	0.16	0.42	0.13	0.12	0.41	0.13	0.12
16	Global Max	0.88	0.18	0.24	0.89	0.18	0.20	0.90	0.18	0.10	0.86	0.18	0.10
17	Global Max	0.77	0.23	0.68	0.76	0.24	0.53	1.14	0.25	0.18	0.57	0.24	0.28
18	Global Max	0.68	0.27	0.52	0.64	0.28	0.38	0.43	0.29	0.26	0.42	0.29	0.26
19	Global Max	0.68	0.18	0.30	0.66	0.18	0.24	0.55	0.18	0.19	0.52	0.18	0.19
20	Global Max	1.02	0.20	0.34	0.95	0.20	0.28	0.91	0.21	0.17	0.92	0.21	0.17
21	Range	0.62	0.22	0.32	0.56	0.21	0.28	0.43	0.20	0.23	0.44	0.21	0.23
22	Rank	0.82	0.11	0.15	0.83	0.10	0.14	0.71	0.10	0.07	0.69	0.10	0.07
23	Global Max	0.71	0.13	0.15	0.71	0.14	0.12	0.54	0.14	0.09	0.54	0.14	0.09
24	Local Max	1.01	0.24	0.38	1.04	0.23	0.26	1.02	0.24	0.13	1.01	0.24	0.13
25	Rank	0.27	0.21	0.10	0.24	0.21	0.09	0.00	0.20	0.09	0.00	0.21	0.09
26	Global Max	0.91	0.16	0.31	0.91	0.17	0.26	0.90	0.17	0.17	0.89	0.17	0.17
27	Local Max	0.45	0.12	0.18	0.45	0.12	0.15	0.11	0.13	0.11	0.13	0.13	0.11
28	Global Max	0.04	0.16	0.12	0.10	0.15	0.11	0.00	0.16	0.09	0.00	0.16	0.09
29	Global Max	0.38	0.19	0.26	0.35	0.19	0.19	0.16	0.20	0.15	0.17	0.20	0.15
30	Global Max	0.67	0.29	0.66	0.67	0.30	0.51	0.35	0.31	0.27	0.34	0.31	0.27
31	Rank	0.32	0.21	0.00	0.32	0.21	0.00	0.30	0.21	0.00	0.30	0.21	0.00
32	Range	0.71	0.19	0.30	0.68	0.20	0.25	0.49	0.19	0.14	0.49	0.20	0.13
33	Rank	0.70	0.20	0.24	0.68	0.20	0.22	0.43	0.20	0.14	0.43	0.20	0.14
34	Local Max	0.20	0.12	0.18	0.21	0.11	0.14	0.00	0.12	0.12	0.00	0.12	0.12
35	Global Max	0.55	0.31	0.36	0.54	0.31	0.26	0.36	0.31	0.20	0.34	0.31	0.20
36	Range	0.80	0.16	0.23	0.80	0.16	0.22	0.71	0.17	0.12	0.64	0.16	0.13
37	Global Max	0.80	0.16	0.29	0.81	0.16	0.20	0.72	0.16	0.12	0.71	0.17	0.12
38	Local Max	0.61	0.22	0.23	0.61	0.21	0.23	0.38	0.22	0.14	0.37	0.22	0.14
39	Global Max	0.01	0.35	0.20	0.00	0.36	0.15	0.00	0.35	0.03	0.00	0.36	0.07
40	Global Max	0.66	0.16	0.35	0.61	0.16	0.23	0.28	0.16	0.15	0.28	0.16	0.15
41	Rank	0.93	0.16	0.21	0.94	0.16	0.19	0.90	0.15	0.12	0.91	0.15	0.12
42	Global Max	0.76	0.17	0.33	0.77	0.18	0.27	0.72	0.20	0.14	0.73	0.20	0.13
43	Local Max	0.80	0.19	0.29	0.79	0.19	0.21	0.69	0.19	0.13	0.68	0.19	0.13
44	Global Max	0.63	0.21	0.29	0.62	0.22	0.25	0.80	0.23	0.13	0.45	0.23	0.18
45	Rank	0.95	0.18	0.47	0.92	0.18	0.41	0.81	0.17	0.25	0.79	0.17	0.25
46	Global Max	0.75	0.16	0.27	0.74	0.17	0.22	0.57	0.17	0.14	0.60	0.16	0.14
47	Range	0.99	0.24	0.29	0.96	0.23	0.25	1.00	0.23	0.15	1.23	0.24	0.13
48	Global Max	0.93	0.22	0.33	0.95	0.22	0.28	0.91	0.23	0.13	0.91	0.23	0.12
49	Local Max	0.44	0.19	0.44	0.49	0.19	0.29	0.92	0.22	0.11	0.28	0.20	0.22
50	Local Max	0.82	0.17	0.30	0.80	0.17	0.25	0.69	0.17	0.18	0.71	0.17	0.18

Appendix B

Figure B.1: Pre-registration of Experiment 1 from Chapter 3.

AS PREDICTED

Author(s)
Anna Trendl (University of Warwick) - a.trendl@warwick.ac.uk
Neil Stewart (University of Warwick) - neil.stewart@wbs.ac.uk
Timothy Mullett (University of Warwick) - t.mullett@warwick.ac.uk

Created: 02/12/2018 03:11 AM (PT)
Public: 04/10/2019 05:54 AM (PT)

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
Can we replicate the attraction effect using real-world, complex stimuli?

3) Describe the key dependent variable(s) specifying how they will be measured.
First, we will ask participants to rate 200 movies on a scale from 1 to 7 and to indicate whether they have seen the movie before. Based on the ratings, we will create choice sets (quadruplets consisting of two triplets) for each participant to test the attraction effect. We will use an LSA solution as well as genre information on the movies to create the choice sets (dissimilar movies with the same rating are competitors; similar movies with at least a 2 point rating difference will be targets/decoys). There will be two types of choice triplets according to a matching criteria (based on the semantic and genre proximity between the movies), strict (where at least one of the semantic/genre criteria are strict) and non-strict matches. The validity of the non-strict choice sets will be checked by asking participants to rate the similarity of the movies after the choice stage. Only participants for whom we can create at least three quadruplets with strict matching will be invited to the choice stage. Our main dependent variable will be the proportion of trials in which the participant chooses the target which was favoured by the decoy out of all trials where participants do not choose a decoy.

4) How many and which conditions will participants be assigned to?
In the choice stage, all participants will face same attraction effect choice situation (target, competitor, decoy movies), but the exact movies will depend on the ratings the participant gave in the screening stage. The movie pairs presented in the similarity rating task will also be specific to the participant.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will test the effect using only the quadruplets selected using the strict criteria, and using all quadruplets, we will also use the half of quadruplets for each subject that show the best similarity scores between target and decoy. We are combining estimation and NHST approaches. For the estimation, we will calculate a 95% confidence interval for the proportion of trials upon which the target movie is chosen (after excluding trials where the decoy was chosen). For the NHST, we will conduct a one-sample t-test to test whether the proportion of trials upon which the target movie is chosen is higher than 0.5 (after excluding trials where the decoy was chosen). Using the trials upon which the decoy was not chosen, we will also run the following mixed effects logistic regression model:
Target_chosen = Read_movie + Target_Competitor_distance + Target_Decoy_distance + Target_Decoy_rating_difference
We aim to run this regression with as full a set of random effects (slopes and intercepts) as the final data allow.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
Based on our pilot data collection (using books as stimuli), we decided to exclude participants who completed the ratings task unusually quickly (fastest 5%), whose ratings were unrealistically skewed (lowest 5% of the entropy distribution) or showed a distinct (including random) time trend (lowest and highest 5% of the autocorrelation distribution). A pilot study analysing the stability of repeated ratings confirmed that people who met any of the above criteria were the ones who were the least consistent repeated ratings (indicating that they indeed did not pay attention to the task).

We will also exclude any trials where the subject selected the decoy.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will collect the data in batches of 50 until we have choice data for 100 participants.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
We ran a pilot study with 33 people (mostly ratings) using books as stimuli. Since most of our subjects were not familiar with the stimuli, we decided to use movies instead but the data was still useful in helping us determine our exclusion criteria. In addition, we are in the process of data collection and was planning to submit this registration before we started data collection, but due to a miscommunication between the researchers it has been slightly delayed. However, we have not looked at our results yet and will not do so until we collected choice data for a 100 participants.

Version of AsPredicted Questions: 2.00

Available at https://aspredicted.org/9j897.pdf
(Permanently archived at http://web.archive.org/web/*/https://aspredicted.org/9j897.pdf)

Wharton University of Pennsylvania CREDIBILITY LAB

Figure B.2: Distribution of the proportion of trials where the target was chosen in Experiment 1 (triplets with strict target-decoy pairs plus the "better half" of the remaining quadruplets). The red dot and error bars show the bootstrapped mean and 95% CIs.

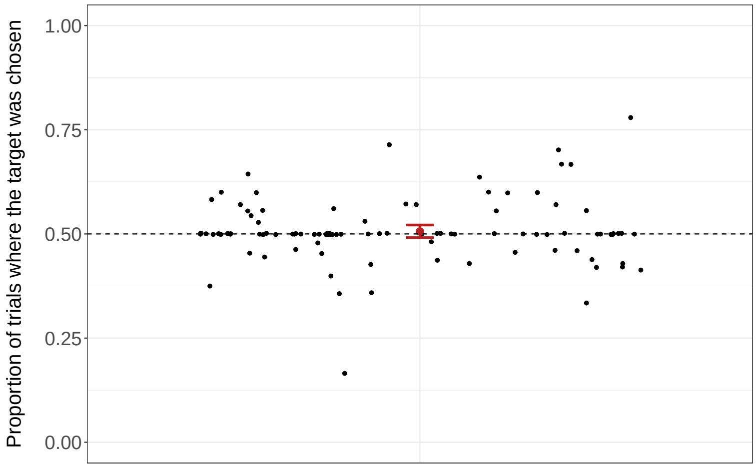


Figure B.3: Pre-registration of Experiment 2 from Chapter 3.

CONFIDENTIAL - FOR PEER-REVIEW ONLY

AsPredicted
Pre-Registration made easy

Attraction in the wild II (#9637)

Created: 04/05/2018 08:45 AM (PT)
Shared: 11/20/2018 02:08 AM (PT)

This pre-registration is not yet public. This anonymized copy (without author names) was created by the author(s) to use during peer-review. A non-anonymized version (containing author names) will become publicly available only if an author makes it public. Until that happens the contents of this pre-registration are confidential.

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
Can we replicate the attraction effect using real-world, complex stimuli? In a previous experiment (#8353), we failed to replicate this effect, and we are interested in whether this is still the case if we use a different method to establish similarity between pairs of stimuli.

3) Describe the key dependent variable(s) specifying how they will be measured.
This is a follow-up experiment on the experiment described in #8353. Changes implemented: we are using 231 movies instead of 200; we are only using the genre information about movies to establish similarity and dissimilarity between them; we will only include participants for whom we can create at least 3 quadruplets; we will ask participants to rate the similarity of each decoy-target and target-competitor pair after the choice phase. The dependent variable is still the proportion of trials in which the participant chooses the target which was favoured by the decoy out of all trials where participants do not choose a decoy.

4) How many and which conditions will participants be assigned to?
In the choice stage, all participants will face same attraction effect choice situation (target, competitor, decoy movies), but the exact movies will depend on the ratings the participant gave in the screening stage. The movie pairs presented in the similarity rating task will also be specific to the participant.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We are combining estimation and NHST approaches. For the estimation, we will calculate a 95% confidence interval for the proportion of trials upon which the target movie is chosen (after excluding trials where the decoy was chosen). For the NHST, we will conduct a one-sample t-test to test whether the proportion of trials upon which the target movie is chosen is higher than 0.5 (after excluding trials where the decoy was chosen). Using the trials upon which the decoy was not chosen, we will also run the following mixed effects logistic regression model:
Target_chosen = Read_movie + Target_Competitor_distance + Target_Decoy_distance + Target_Decoy_rating_difference
We aim to run this regression with as full a set of random effects (slopes and intercepts) as the final data allow.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will exclude participants who completed the ratings task unusually quickly (fastest 5%), whose ratings were unrealistically skewed (lowest 5% of the entropy distribution) or showed a distinct (including random) time trend (lowest and highest 5% of the autocorrelation distribution). We will exclude participants for whom we cannot create at least 3 quadruplets. We will also exclude any trials where the subject selected the decoy.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will collect the data in batches of 50 until we have choice data for 100 participants (after the exclusion criteria have been applied).

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
We have so far collected similarity ratings from 60 people for pairs of movies (207 each) to select the stimuli for this experiment. Based on this data, we are using 253 movie pairs in this experiment (these are the movies that had a mean similarity rating of 4.5 or above). We have not collected any choice data yet.

Verify authenticity:http://aspredicted.org/blind.php?x=iw8iu9

Version of AsPredicted Questions: 2.00

Appendix C

Figure C.1: Pre-registration of Experiment 1 from Chapter 4.

AsPredicted
Pre-Registration made easy

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Teapot experiment I (#12526)

Created: 07/10/2018 06:32 AM (PT)
Shared: 09/13/2018 03:58 AM (PT)

This pre-registration is not yet public. This anonymized copy (without author names) was created by the author(s) to use during peer-review. A non-anonymized version (containing author names) will become publicly available only if an author makes it public. Until that happens the contents of this pre-registration are confidential.

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Does the attraction effect go away when the two attribute dimensions cannot be processed independently?

3) Describe the key dependent variable(s) specifying how they will be measured.
The main dependent variable is the proportion of trials in each condition (numerical, pictorial) on which the target was chosen (excluding trials where the decoy was chosen).

4) How many and which conditions will participants be assigned to?
There are two conditions: numerical vs pictorial representation of stimuli. Each participant will complete both conditions, but the order of the conditions will be randomised.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
First we are going to test if the order of the conditions has an effect on the strength of the attraction effect. This will be tested with a mixed effects regression where we will be including an interaction between condition and order. We do not expect order to have an effect, but if it does, we will only analyse each participant's first condition. Then, we will assess the strength of the effect by condition by calculating a 95% confidence interval for our key variable in both conditions. Using a one-sided t-test, we will test whether the mean of our key variable is significantly greater than 0.5 in the two conditions. We will also test the difference between these two means with a paired/unpaired (depending on whether order will have an effect) sample t-test. Supplementary analyses will involve testing the effect of the number of practice trials on the strength of the effect (mixed effects model), and the effect of the target-decoy value difference on the likelihood of choosing he target (mixed effects logistic regression).

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will exclude partcipants whose 1) accuracy on the catch trials are 2.5 standard deviations below the average 2) choice pattern is in the the lowest 2.5% of the entropy distribution 3) autocorrelation is in the upper or lower 2.5% of the autocorrelation distribution. We will exclude 2.5% of fastest trials. We will also exclude any trials where the subject selected the decoy.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will collect data from a 100 participants.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Verify authenticity:http://aspredicted.org/blind.php?x=p7tg96

Version of AsPredicted Questions: 2.00

Table C.1: Odds-ratios and 95% CIs from a mixed-effects logistic model with subject-specific intercepts, Experiment 2. (T – Target, C – Competitor, D – Decoy)

	Dependent variable:
	Target chosen proportion
Condition:Numerical	1.028 (0.897, 1.179)
Order:Pictorial first	0.974 (0.811, 1.172)
Practice attempts	0.942** (0.891, 0.997)
Condition:NumericalXOrder:Pictorial first	1.351*** (1.124, 1.625)
Constant	1.038 (0.900, 1.195)
Observations	144
Log Likelihood	−505.947
Akaike Inf. Crit.	1,023.895
Bayesian Inf. Crit.	1,041.713
Note:	*p<0.1; **p<0.05; ***p<0.01

References

Ahn, S., Kim, J., & Ha, Y. W. (2015). Feedback weakens the attraction effect in repeated choices. *Marketing Letters*, *26*(4), 449–459.

Akritas, A. G., & Malaschonok, G. I. (2004). Applications of singular-value decomposition (SVD). *Mathematics and Computers in Simulation*, *67*(1-2), 15–31.

Alcohol Concern. (2014). *Alcohol and football.* Retrieved from `www.alcoholconcern.org.uk`

Angner, E., & Loewenstein, G. (2012). Behavioral Economics. In U. Mäki (Ed.), *Handbook of the philosophy of science: Philosophy of economics* (pp. 641–690). Amsterdam, The Netherlands: Elsevier Publishing.

Anton, H., & Rorres, C. (2014). *Elementary Linear Algebra: Applications version* (11th ed.). Hoboken, NJ: John Wiley & Sons.

Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, *3*(5), 396–403.

Arrow, K. J., & Debreu, G. (1954). Existence of an Equilibrium for a Competitive Economy. *Econometrica*, *22*(3), 265–290.

Ashraf, N., Camerer, C. F., & Loewenstein, G. (2005). Adam Smith, Behavioral Economist. *Journal of Economic Perspectives*, *19*(3), 131–145.

Attneave, F. (1950). Dimensions of Similarity. *The American Journal of Psychology*, *63*(4), 516–556.

Barton, R. R., & Ivey, J. S. (1996). Nelder-Mead Simplex Modifications for Simulation Optimization. *Management Science*, *42*(7), 954–973.

Bateson, M., Healy, S. D., & Hurly, T. A. (2002). Irrational choices in hummingbird foraging behaviour. *Animal Behaviour*, *63*(3), 587–596.

BBC Sport. (2018, July 16). *World Cup 2018: Record-breaking online & viewing figures for BBC.* Retrieved from https://www.bbc.co.uk/sport/football/44850988

Bellis, M. A., Leckenby, N., Hughes, K., Luke, C., Wyke, S., & Quigg, Z. (2012). Nighttime assaults: using a national emergency department monitoring system to predict occurrence, target prevention and plan services. *BMC Public Health*, *12*(746).

Berkowitsch, N. A. J., Scheibehenne, B., & Rieskamp, J. (2014). Rigorously Testing Multialternative Decision Field Theory Against Random Utility Models. *Journal of Experimental Psychology: General*, *143*(3), 1331–1348.

Bhatia, S. (2013). Associations and the Accumulation of Preference. *Psychological Review*, *120*(3), 522–543.

Bhatia, S., & Stewart, N. (2018). Naturalistic multiattribute choice. *Cognition*, *179*, 71–88.

Bigelow, J. L., Edwards, A., & Edwards, L. (2016). Detecting Cyberbullying using Latent Semantic Indexing. In *Proceedings of the first international workshop on computational methods for cybersafety* (pp. 11–14). Indianapolis, IN, USA: ACM.

Bird, G. D., Lauwereyns, J., & Crawford, M. T. (2012). The role of eye movements in decision making and the prospect of exposure effects. *Vision Research*, *60*, 16–21.

Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, *130*, 99–113.

Boyce, C. J., Brown, G. D., & Moore, S. C. (2010). Money and Happiness: Rank of Income, Not Income, Affects Life Satisfaction. *Psychological Science*, *21*(4), 471–475.

Brimicombe, A., & Cafe, R. (2012). Beware, Win or Lose: Domestic Violence and the World Cup. *Significance*, *9*(5), 32–35.

Brooks-Hay, O., & Lombard, N. (2018). 'Home game': domestic abuse and football. *Journal of Gender-Based Violence*, *2*(1), 93–108.

Brown, G. D., Gardner, J., Oswald, A. J., & Qian, J. (2008). Does Wage Rank Affect Employees' Well-being? *Industrial Relations*, *47*(3), 355–389.

Bruni, L., & Guala, F. (2001). Vilfredo Pareto and the Epistemological Foundations of Choice Theory. *History of Political Economy*, *33*(1), 21–49.

Bruni, L., & Luigino. (2013). The Paretian Turn – The foundation of the theory of rational choice, and its discontents. *European Journal of Social Sciences*(51-2), 47–64.

Bruni, L., & Sugden, R. (2007). The road not taken: How psychology was removed from economics, and how it might be brought back. *Economic Journal*, *117*(516), 146–173.

Burns, D. M. (2014). *The Many Faces of Garner Interference* (Doctoral dissertation, Indiana University). Retrieved from `https://scholarworks.iu.edu/dspace/handle/2022/18756`

Busemeyer, J. R., Gluth, S., Rieskamp, J., & Turner, B. M. (2019). Cognitive and Neural Bases of Multi-Attribute, Multi-Alternative, Value-based Decisions. *Trends in Cognitive Sciences*, *23*(3), 251–263.

Callaghan, J. E. M., Alexander, J. H., Sixsmith, J., & Fellin, L. C. (2018). Beyond "Witnessing": Children's Experiences of Coercive Control in Domestic Violence and Abuse. *Journal of Interpersonal Violence*, *33*(10), 1551–1581.

Capaldi, D. M., Knoble, N. B., Shortt, J. W., & Kim, H. K. (2012). A Systematic Review of Risk Factors for Intimate Partner Violence. *Partner Abuse*, *3*(2), 231–280.

Card, D., & Dahl, G. B. (2011). Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior. *The Quarterly Journal of Economics*, *126*(1), 103–143.

Cataldo, A. M., & Cohen, A. L. (2018). The comparison process as an account of variation in the attraction, compromise, and similarity effects. *Psychonomic Bulletin & Review*, *26*(3), 934–942.

Chang, K. H. (2012). Stochastic Nelder-Mead simplex method - A new globally convergent direct search method for simulation optimization. *European Journal of Operational Research*, *220*(3), 684–694.

Choplin, J. M., & Hummel, J. E. (2005). Comparison-induced decoy effects. *Memory & Cognition*, *33*(2), 332–343.

Chung, H.-K., Sjöström, T., Lee, H.-J., Lu, Y.-T., Tsuo, F.-Y., Chen, T.-S., ... Huang, C.-Y. (2017). Why Do Irrelevant Alternatives Matter? An fMRI-TMS Study of Context-Dependent Preferences. *The Journal of Neuroscience*, *37*(48), 11647–11661.

Churchland, A. K., Kiani, R., Chaudhuri, R., Wang, X. J., Pouget, A., & Shadlen, M. N. (2011). Variance as a Signature of Neural Computations during Decision Making. *Neuron*, *69*(4), 818–831.

Clark, A. E., Frijters, P., & Shields, M. A. (2008). Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature*, *46*(1), 95–144.

Cohen, P. M., & Santos, L. R. (2017). Capuchins (Cebus apella) fail to show an asymmetric dominance effect. *Animal Cognition*, *20*(2), 331–345.

Conover, M. D., Gonçalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011). Predicting the Political Alignment of Twitter Users. In *2011 ieee international conference on privacy, security, risk and trust and ieee international conference on on social computing* (pp. 192–199). Boston, MA, USA: IEEE.

Costa, B. M., Kaestle, C. E., Walker, A., Curtis, A., Day, A., Toumbourou,

J. W., & Miller, P. (2015). Longitudinal predictors of domestic violence perpetration and victimization: A systematic review. *Aggression and Violent Behavior*, *24*, 261–272.

Daly, M., Boyce, C., & Wood, A. (2015). A Social Rank Explanation of How Money Influences Health. *Health Psychology*, *34*(3), 222–230.

Davies, G. (2018, December 30). England's unexpected World Cup progression behind record-high alcohol poisoning cases, NHS says. *The Telegraph*. Retrieved from https://www.telegraph.co.uk/news/2018/12/30/englands-unexpected-world-cup-progression-behind-record-high/

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407.

de Haan, T., & van Veldhuizen, R. (2015). Willpower depletion and framing effects. *Journal of Economic Behavior & Organization*, *117*, 47–61.

Demeritt, A. D., & Hoff, K. (2018). *The Making of Behavioral Development Economics* (Tech. Rep. No. 5694). The World Bank. Retrieved from http://documents.worldbank.org/curated/en/708521516820237922/The-making-of-behavioral-development-economics

Dickson, A., Jennings, C., & Koop, G. (2016). Domestic Violence and Football in Glasgow: Are Reference Points Relevant? *Oxford Bulletin of Economics and Statistics*, *78*(1), 1–21.

Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, *47*(3), 304–322.

Dixon, K. (2014). The football fan and the pub: An enduring relationship. *International Review for the Sociology of Sport*, *49*(3-4), 382–399.

Doyle, J. R., O'Connor, D. J., Reynolds, G. M., & Bottomley, P. A. (1999). The robustness of the asymmetrically dominated effect: Buying frames, phantom alternatives, and in-store purchases. *Psychology & Marketing*, *16*(3), 225–243.

Drakopoulos, S. A., & Katselidis, I. (2017). *The Relationship between Psychology and Economics: Insights from the History of Economic Thought* (MPRA Working Paper No. 77485). Retrieved from https://mpra.ub.uni-muenchen.de/77485/

Edwards, S. C., & Pratt, S. C. (2009). Rationality in collective decision-making by ant colonies. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1673), 3655–3661.

Ending Domestic Abuse - Getting it right first time (Policy Report). (2015). Safe Lives. Retrieved from http://www.safelives.org.uk/policy-evidence/getting-it-right-first-time

Eriksen, C. W., & Hake, H. W. (1955). Multidimensional stimulus differences and accuracy of discrimination. *Journal of Experimental Psychology*, *50*(3), 153–160.

Fahmy, E., & Williamson, E. (2018). Poverty and domestic violence and abuse (DVA) in the UK. *Journal of Gender-Based Violence*, *2*(3), 481–501.

Farmer, G. D., El-Deredy, W., Howes, A., & Warren, P. (2015). The attraction effect in motor planning decisions. *Judgment and Decision Making*, *10*(5), 503–510.

Farmer, G. D., Warren, P. A., El-Deredy, W., & Howes, A. (2017). The Effect of Expected Value on Attraction Effect Preference Reversals. *Journal of Behavioral Decision Making*, *30*(4), 785–793.

Fehr, E., & Rangel, A. (2011). Neuroeconomic Foundations of Economic Choice—Recent Advances. *Journal of Economic Perspectives*, *25*(4), 3–30.

Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, *114*(3), 817–868.

Ferrari, G., Agnew-Davies, R., Bailey, J., Howard, L., Howarth, E., Peters, T. J.,

... Feder, G. S. (2014). Domestic violence and mental health: a cross-sectional survey of women seeking help from domestic violence support services. *Global Health Action*, *7*(1), 25519.

Fisher, G. (2017). An attentional drift diffusion model over binary-attribute choice. *Cognition*, *168*, 34–45.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annual Review of Psychology*, *67*, 641–666.

Fraser McKevitt. (2018, July 24). *Sizzling summer lights up the grocery market.* Kantar Worldpanel. Retrieved from `https://www.kantarworldpanel.com/global/News/Sizzling-summer-lights-up-the-grocery-market-in-UK`

Frederick, S., Lee, L., & Baskin, E. (2014). The Limits of Attraction. *Journal of Marketing Research*, *51*(4), 487–507.

Friedman, M. (1953). The Methodology of Positive Economics. In *Essays in positive economics* (pp. 3–43). Chicago, Illinois: The University of Chicago Press.

Gantz, W., Bradley, S. D., & Wang, Z. (2006). Televised NFL Games, the Family, and Domestic Violence. In A. A. Raney & J. Bryant (Eds.), *Handbook of sports and media* (pp. 398–414). Taylor & Francis. Retrieved from `https://www.taylorfrancis.com/books/9780203873670`

Garcia Coll, C., Bearer, E. L., Lerner, R. M., Bearer, E. L., & Lerner, R. M. (2004). *Nature and Nurture: The Complex Interplay of Genetic and Environmental Influences on Human Behavior and Development.* Mahwah, NJ: Erlbaum.

Garner, W., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*(3), 225–241.

Garner, W., & Lee, W. (1962). An Analysis of Redundancy in Perceptual

Discrimination. *Perceptual and Motor Skills*, *15*(2), 367–388.

Gigerenzer, G. (2016). Towards a Rational Theory of Heuristics. In R. Frantz & L. Marsh (Eds.), *Minds, models and milieux: Commemorating the centennial of the birth of herbert simon* (pp. 34–59). New York: Palgrave Macmillan.

Glaholt, M. G., & Reingold, E. M. (2009). Stimulus exposure and gaze bias: A further test of the gaze cascade model. *Attention, Perception, & Psychophysics*, *71*(3), 445–450.

Gluth, S., Hotaling, J. M., & Rieskamp, J. (2017). The Attraction Effect Modulates Reward Prediction Errors and Intertemporal Choices. *The Journal of Neuroscience*, *37*(2), 371–382.

Gornall, J. (2014). World Cup 2014: festival of football or alcohol? *BMJ*, *348*, g3772.

Graham, A., & Adams, J. (2013). Alcohol marketing in televised English professional football: a frequency analysis. *The Lancet*, *382*(S4), 343–348.

Gwinn, R., Leber, A. B., & Krajbich, I. (2019). The spillover effects of attentional learning on value-based choice. *Cognition*, *182*, 294–306.

Hedgcock, W., & Rao, A. R. (2009). Trade-Off Aversion as an Explanation for the Attraction Effect: A Functional Magnetic Resonance Imaging Study. *Journal of Marketing Research*, *46*(1), 1–13.

Hergenhahn, B. R., & Henley, T. (2013). *An Introduction to the History of Psychology* (7th ed.). Belmont, CA, USA: Wadsworth Publishing Co.

Highhouse, S. (1996). Context-Dependent Selection: The Effects of Decoy and Phantom Job Candidates. *Organizational Behavior and Human Decision Processes*, *65*(1), 68–76.

Home Office. (2012). *Cross-government definition of domestic violence consultation* (Consultation outcome). Retrieved from `https://www.gov.uk/government/consultations/cross-government-definition-of-domestic-violence-consultation`

Home Office. (2016). *Strategy to end violence against women and girls: 2016 to 2020* (Policy paper). Retrieved from https://www.gov.uk/government/publications/strategy-to-end-violence-against-women-and-girls-2016-to-2020

Home Office. (2018). *Police funding for England and Wales 2015 to 2019* (Official Statistics). Retrieved from https://www.gov.uk/government/statistics/police-funding-for-england-and-wales-2015-to-2019

Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2013). Multidimensional scaling. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(1), 93–103.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research*, *9*(1), 90–98.

Huber, J., Payne, J. W., & Puto, C. P. (2014). Let's Be Honest About the Attraction Effect. *Journal of Marketing Research*, *51*(4), 520–525.

Ingre, M., & Nilsonne, G. (2018). Estimating statistical power, posterior probability and publication bias of psychological research using the observed replication rate. *Royal Society Open Science*, *5*(9), 181190.

Ipsos MORI. (2003, December 22). *Rugby Union 'Britain's Second Most Popular Sport'.* Press Release. Retrieved from https://www.ipsos.com/ipsos-mori/en-uk/rugby-union-britains-second-most-popular-sport

Javaid, A. (2015). The role of alcohol in intimate partner violence: Causal behaviour or excusing behaviour? *British Journal of Community Justice*, *13*(1), 75–92.

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review*, *93*(5), 1449–1475.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–292.

Khalifeh, H., Hargreaves, J., Howard, L. M., & Birdthistle, I. (2013). Intimate Partner Violence and Socioeconomic Deprivation in England: Findings From a National Cross-Sectional Survey. *American Journal of Public Health*, *103*(3), 462–472.

Kidd, B. (2013). Sports and masculinity. *Sport in Society*, *16*(4), 553–564.

Kim, S., & Hasher, L. (2005). The Attraction Effect in Decision Making: Superior Performance by Older Adults. *Quarterly Journal of Experimental Psychology Section A*, *58*(1), 120–133.

Kirby, S., Francis, B., & O'Flaherty, R. (2014). Can the FIFA World Cup Football (Soccer) Tournament Be Associated with an Increase in Domestic Abuse? *Journal of Research in Crime and Delinquency*, *51*(3), 259–276.

Krajbich, I. (2019). Accounting for attention in sequential sampling models of decision making. *Current Opinion in Psychology*, *29*, 6–11.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Krajbich, I., & Dean, M. (2015). How can neuroscience inform economics? *Current Opinion in Behavioral Sciences*, *5*, 51–57.

Krajbich, I., Oud, B., & Fehr, E. (2014). Benefits of Neuroeconomic Modeling: New Policy Interventions and Predictors of Preference. *American Economic Review*, *104*(5), 501–506.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13852–13857.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1–27.

Latty, T., & Beekman, M. (2011). Irrational decision-making in an amoeboid organism: transitivity and context-dependent preferences. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1703), 307–312.

Leahey, T. H. (1979). Something old, something new: Attention in Wundt and modern cognitive psychology. *Journal of the History of the Behavioral Sciences*, *15*(3), 242–252.

Leite, F. P., & Ratcliff, R. (2011). What cognitive processes drive response biases? A diffusion model analysis. *Judgment and Decision Making*, *6*(7), 651–687.

Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). *Mining of Massive Datasets* (2nd ed.). Cambridge, UK: Cambridge University Press.

Levitt, S. D., & List, J. A. (2007). Viewpoint: On the generalizability of lab behaviour to the field. *Canadian Journal of Economics*, *40*(2), 347–370.

Liam Bannon. (2016). *Norfolk Domestic Abuse Needs Assessment* (Health Needs Assessments). Office of the Police and Crime Commissioner for Norfolk. Retrieved from www.norfolkinsight.org.uk/resource/view?resourceId=1733

Liew, S. X., Howe, P. D. L., & Little, D. R. (2016). The appropriacy of averaging in the study of context effects. *Psychonomic Bulletin & Review*, *23*(5), 1639–1646.

Lim, S.-L., O'Doherty, J. P., & Rangel, A. (2011). The Decision Value Computations in the vmPFC and Striatum Use a Relative Value Code That is Guided by Visual Attention. *Journal of Neuroscience*, *31*(37), 13214–13223.

Little, D. R., Nosofsky, R. M., & Denton, S. E. (2011). Response-time tests of logical-rule models of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 1–27.

Little, D. R., Nosofsky, R. M., Donkin, C., & Denton, S. E. (2013). Logical rules and the classification of integral-dimension stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 801–820.

Lockhead, G. R. (1966). Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology*, *72*(1), 95–104.

Lockhead, G. R., & Hinson, J. (1986). Range and sequence effects in judgment.

Perception & Psychophysics, *40*(1), 53–61.

Loomes, G., & Sugden, R. (1982). Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal*, *92*(368), 805–824.

Louie, K., Khaw, M. W., & Glimcher, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences*, *110*(15), 6139–6144.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis.* New York: Wiley.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, *15*(3), 215–233.

Malkoc, S. A., Hedgcock, W., & Hoeffler, S. (2013). Between a rock and a hard place: The failure of the attraction effect among unattractive alternatives. *Journal of Consumer Psychology*, *23*(3), 317–329.

Mandler, G. (2007). *A History of Modern Experimental Psychology.* Cambridge, MA: The MIT Press.

Mao, W., & Oppewal, H. (2012). The attraction effect is more pronounced for consumers who rely on intuitive reasoning. *Marketing Letters*, *23*(1), 339–351.

Maylor, E. A., & Roberts, M. A. (2007). Similarity and attraction effects in episodic memory judgments. *Cognition*, *105*(3), 715–723.

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144.

Mohr, P. N., Heekeren, H. R., & Rieskamp, J. (2017). Attraction Effect in Risky Choice Can Be Explained by Subjective Distance between Choice Alternatives. *Scientific Reports*, *7*(8942).

Mueller, C. J., & Kuchinke, L. (2016). Individual differences in emotion word processing: A diffusion model analysis. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(3), 489–501.

Mulder, M. J., van Maanen, L., & Forstmann, B. U. (2014). Perceptual decision neurosciences - a model-based review. *Neuroscience*, *277*, 872–884.

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Mullett, T. L., & Stewart, N. (2016). Implications of Visual Attention Phenomena for Models of Preferential Choice. *Decision*, *3*(4), 231–253.

Mullett, T. L., & Tunney, R. J. (2013). Value representations by rank order in a distributed network of varying context dependency. *Brain and Cognition*, *82*(1), 76–83.

National Center for Domestic Violence. (2018). *The Not-So-Beautiful Game.* Retrieved from `https://www.ncdv.org.uk/the-not-so-beautiful-game`

Neumann, J. V., & Morgenstern, O. (1947). *Theory of Games and Economic Behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

Newhall, S. M. (1940). Preliminary Report of the O.S.A. Subcommittee on the Spacing of the Munsell Colors. *Journal of the Optical Society of America*, *30*(12), 617–645.

Niedrich, R. W., Weathers, D., Hill, R. C., & Bell, D. R. (2009). Specifying Price Judgments with Range–Frequency Theory in Models of Brand Choice. *Journal of Marketing Research*, *46*(5), 693–702.

Noguchi, T., & Stewart, N. (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition*, *132*(1), 44–56.

Office for National Statistics. (2017). *Labour Market Profile - West Midlands (Met County).* Retrieved from `https://www.nomisweb.co.uk/reports/lmp/la/1967128614/report.aspx{#}tabrespop`

Office for National Statistics. (2018, November 22). *Domestic abuse in England and Wales: year ending March 2018* (Statistical Bulletin). Retrieved from `https://www.gov.uk/government/statistics/police-funding-for-england-and-wales-2015-to-2019`

Office for National Statistics. (2019, February 7). *Homicide in England and Wales: year ending March 2018* (Statistical Bulletin). Retrieved from https://www.ons.gov.uk/releases/homicideinenglandandwalesyearendingmarch2018

Oliver, A. (2013). From Nudging to Budging: Using Behavioural Economics to Inform Public Sector Policy. *Journal of Social Policy*, *42*(4), 685–700.

Oliver, R., Alexander, B., Roe, S., & Wlasny, M. (2019, January 21). *The economic and social costs of domestic abuse* (Home Office Research Report 107). Retrieved from https://www.gov.uk/government/publications/the-economic-and-social-costs-of-domestic-abuse

Oths, K., & Robertson, T. (2007). Give Me Shelter: Temporal Patterns of Women Fleeing Domestic Abuse. *Human Organization*, *66*(3), 249–260.

Padoa-Schioppa, C. (2009). Range-Adapting Representation of Economic Value in the Orbitofrontal Cortex. *Journal of Neuroscience*, *29*(44), 14004–14014.

Padoa-Schioppa, C., & Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature Neuroscience*, *11*, 95–102.

Padoa-Schioppa, C., & Conen, K. E. (2017). *Orbitofrontal Cortex: A Neural Circuit for Economic Decisions* (Vol. 96) (No. 4).

Parducci, A. (1963). Range-frequency compromise in judgment. *Psychological Monographs: General and Applied*, *77*(2), 1–50.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, *72*(6), 407–418.

Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M. J., & Richardson, D. C. (2015). Biasing moral decisions by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences*, *112*(13), 4170–4175.

Parrish, A. E., Afrifa, E., & Beran, M. J. (2018). Exploring decoy effects on computerized task preferences in rhesus monkeys (Macaca mulatta). *Animal Behavior and Cognition*, *5*(2), 235–253.

Parrish, A. E., Evans, T. A., & Beran, M. J. (2015). Rhesus macaques (Macaca mulatta) exhibit the decoy effect in a perceptual discrimination task. *Attention, Perception, and Psychophysics*, *77*(5), 1715–1725.

Parry, K. D., Jones, I., & Wann, D. L. (2014). An Examination of Sport Fandom in the United Kingdom: A Comparative Analysis of Fan Behaviors, Socialization Processes, and Team Identification. *Journal of Sport Behavior*, *37*(3), 251–267.

Peralta, R. L., & Tuttle, L. A. (2013). Male Perpetrators of Heterosexual-Partner-Violence: The Role of Threats to Masculinity. *The Journal of Men's Studies*, *21*(3), 255–276.

Peralta, R. L., Tuttle, L. A., & Steele, J. L. (2010). At the Intersection of Interpersonal Violence, Masculinity, and Alcohol Use: The Experiences of Heterosexual Male Perpetrators of Intimate Partner Violence. *Violence Against Women*, *16*(4), 387–409.

Pettibone, J. C. (2012). Testing the effect of time pressure on asymmetric dominance and compromise decoys in choice. *Judgment and Decision Making*, *7*(4), 513–523.

Pinker, S. (2004). Why nature & nurture won't go away. *Daedalus*, *133*(4), 5–17.

Pocheptsova, A., Amir, O., Dhar, R., & Baumeister, R. F. (2009). Deciding without Resources: Resource Depletion and Choice in Context. *Journal of Marketing Research*, *46*(3), 344–355.

Prpic, M., & Rosamund, S. (2018). *Violence against women in the EU: State of play* (Briefing Paper Series No. 630296). Retrieved from `http://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS{_}BRI(2018)630296`

Quigg, Z., Hughes, K., & Bellis, M. A. (2013). Effects of the 2010 World

Cup football tournament on emergency department assault attendances in England. *The European Journal of Public Health*, *23*(3), 383–385.

Rangel, A., & Clithero, J. A. (2012). Value normalization in decision making: theory and evidence. *Current Opinion in Neurobiology*, *22*(6), 970–981.

Ratcliff, R. (1978). A Theory of Memory Retrieval. *Psychological Review*, *85*(2), 59–108.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review*, *111*(1), 159–182.

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922.

Ratcliff, R., McKoon, G., & van Zandt, T. (1999). Connectionist and Diffusion Models of Reaction Time. *Psychological Review*, *106*(2), 261–300.

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, *9*(5), 347–356.

Reb, J., Li, A., & Bagger, J. (2018). Decoy effect, anticipated regret, and preferences for work–family benefits. *Journal of Occupational and Organizational Psychology*, *91*(3), 441–464.

Rieber, R. W., & Robinson, D. K. (Eds.). (2001). *Wilhelm Wundt in History – The Making of a Scientific Psychology.* New York: Springer Science & Business Media.

Riley, J. (2008). Utilitarianism and Economic Theory. In *The New Palgrave Dictionary of Economics.* London: Palgrave Macmillan.

Riordan, C. (2018, August 7). *Using football as a scapegoat for domestic abuse "trivialises problem".* The Sunday Times. Retrieved from `https://www.thetimes.co.uk/article/using-football-as-a-scapegoat-for-domestic-abuse-trivialises-problem-t932sn9ql`

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative Decision Field Theory: A Dynamic Connectionist Model of Decision Making.

Psychological Review, *108*(2), 370–392.

Sabo, D., Gray, P. M., & Moore, L. A. (2000). Domestic Violence and Televised Athletic Events:'It's a Man Thing'. In Jim McKay; Michael A. Messner; Don Sabo (Ed.), *Masculinities, gender relations, and sport* (pp. 127–146). Thousand Oaks: SAGE Publications, Inc.

Sachs, C. J., & Chu, L. D. (2000). The Association Between Professional Football Games and Domestic Violence in Los Angeles County. *Journal of Interpersonal Violence*, *15*(11), 1192–1201.

Schwartz, J. A., & Chapman, G. B. (1999). Are More Options Always Better? The Attraction Effect in Physicians' Decisions about Medications. *Medical Decision Making*, *19*(3), 315–323.

Scottish Government. (2018, November 27). *Domestic Abuse in Scotland: 2017-2018 statistics* (Statistical Bulletin). Retrieved from `https://www.gov.scot/publications/domestic-abuse-recorded-police-scotland-2017-18/`

Sedikides, C., Ariely, D., & Olsen, N. (1999). Contextual and Procedural Determinants of Partner Selection: Of Asymmetric Dominance and Prominence. *Social Cognition*, *17*(2), 118–139.

Sen, A. (1977). Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philosophy & Public Affairs*, *6*(4), 317–344.

Sent, E.-M. (2004). Behavioral Economics: How Psychology Made Its (Limited) Way Back Into Economics. *History of Political Economy*, *36*(4), 735–760.

Sent, E.-M. (2005). Simplifying Herbert Simon. *History of Political Economy*, *37*(2), 227–232.

Shadlen, M. N., & Newsome, W. T. (2001). Neural Basis of a Perceptual Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey. *Journal of Neurophysiology*, *86*(4), 1916–1936.

Shafir, S., Waite, T. A., & Smith, B. H. (2002). Context-dependent violations of rational choice in honeybees (Apis mellifera) and gray jays (Perisoreus canadensis). *Behavioral Ecology and Sociobiology*, *51*(2), 180–187.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*(1), 54–87.

Shepard, R. N. (1980). Multidimensional Scaling, Tree-Fitting, and Clustering. *Science*, *210*(4468), 390–398.

Shepp, B. E., & Ballesteros, S. (1989). *Object Perception: Structure and Process.* New York: Psychology Press.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*, 1317–1322.

Shipman, J. W. (2012). *Introduction to color theory.* New Mexico Institute of Mining and Technology. Retrieved from `http://infohost.nmt.edu/{~}shipman/doc/colortheory`

Simonson, I. (1989). Choice Based on Reasons: The Case of Attraction and Compromise Effects. *Journal of Consumer Research*, *16*(2), 158–174.

Simonson, I. (2014). Vices and Virtues of Misguided Replications: The Case of Asymmetric Dominance. *Journal of Marketing Research*, *51*(4), 514–519.

Sivakumar, K. (2016). A unified conceptualization of the attraction effect. *AMS Review*, *6*(1-2), 39–58.

Smith, R. (2013). *Between Mind and Nature - A History of Psychology.* London: Reaktion Books Ltd.

Smith, S. M., & Krajbich, I. (2019). Gaze Amplifies Value in Decision Making. *Psychological Science*, *30*(1), 116–128.

Soltani, A., de Martino, B., & Camerer, C. (2012). A Range-Normalization Model of Context-Dependent Choice: A New Model and Evidence. *PLoS Computational Biology*, *8*(7), e1002607.

Sue O'Curry, Y. P., & Pitts, R. (1995). The Attraction Effect and Political Choice in Two Elections. *Journal of Consumer Psychology*, *4*(1), 85–101.

Swallow, J. (2017). *An exploratory study of women's experiences regarding the interplay between domestic violence and abuse and sports events* (Doctoral dissertation, University of Chester). Retrieved from `https://chesterrep.openrepository.com/handle/10034/620835`

Teodorescu, A. R., Moran, R., & Usher, M. (2016). Absolutely relative or relatively absolute: violations of value invariance in human decision making. *Psychonomic Bulletin and Review*, *23*(1), 22–38.

Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychological Review*, *120*(1), 1–38.

Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, *8*(3), 201–207.

Torgerson, W. S. (1958). *Theory and Methods of Scaling.* New York: Wiley.

Towers, J. (2015). Making the links between economic inequality and intimate partner violence. *Safe – The Domestic Abuse Quarterly*, *53*, 22–25.

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The Multiattribute Linear Ballistic Accumulator Model of Context Effects in Multialternative Choice. *Psychological Review*, *121*(2), 179–205.

Trueblood, J. S., Brown, S. D., Heathcote, A., & Busemeyer, J. R. (2013). Not Just for Consumers: Context Effects Are Fundamental to Decision Making. *Psychological Science*, *24*(6), 901–908.

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences*, *109*(24), 9659–9664.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*(4), 281–299.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.

Usher, M., & McClelland, J. L. (2001). The Time Course of Perceptual Choice: The Leaky, Competing Accumulator Model. *Psychological Review*, *108*(3), 550–592.

Usher, M., & McClelland, J. L. (2004). Loss Aversion and Inhibition in Dynamical Models of Multialternative Choice. *Psychological Review*, *111*(3), 757–769.

van der Meij, L., Almela, M., Hidalgo, V., Villada, C., IJzerman, H., van Lange, P. A. M., & Salvador, A. (2012). Testosterone and Cortisol Release among Spanish Soccer Fans Watching the 2010 World Cup Final. *PLoS ONE*, *7*(4), e34814.

Vincent, J., & Harris, J. (2014). 'They think it's all Dover!' Popular newspaper narratives and images about the English football team and (re)presentations of national identity during Euro 2012. *Soccer and Society*, *15*(2), 222–240.

Vincent, J., Kian, E. M., Pedersen, P. M., Kuntz, A., & Hill, J. S. (2010). England expects: English newspapers' narratives about the English football team in the 2006 World Cup. *International Review for the Sociology of Sport*, *45*(2), 199–223.

Vlaev, I., & Chater, N. (2006). Game relativity: How context influences strategic decision making. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(1), 131–149.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory and Cognition*, *32*(7), 1206–1220.

Walby, S., & Allen, J. (2004, March). *Domestic violence, sexual assault and stalking: Findings from the British Crime Survey* (Home Office Research Study 276). Retrieved from `http://openaccess.city.ac.uk/21697/`

Walby, S., Towers, J., & Francis, B. (2016). Is Violent Crime Increasing or Decreasing? a New Methodology to Measure Repeat Attacks Making Visible the Significance of Gender and Domestic Relations. *British Journal of Criminology*, *56*(6), 1203–1234.

Watkinson, P., Wood, A. M., Lloyd, D. M., & Brown, G. D. (2013). Pain ratings reflect cognitive context: A range frequency model of pain perception. *Pain*, *154*(5), 743–749.

White, G. F., Katz, J., & Scarborough, K. E. (1992). The Impact of Professional Football Games Upon Violent Assaults on Women. *Violence and Victims*, *7*(2), 157–171.

Williams, D. J., & Neville, F. G. (2014). Sport-related domestic violence: exploring the complex relationship between sporting events and domestic violence. In Myra F. Taylor; Julie Ann Pooley; Robert S. Taylor (Ed.), *Overcoming domestic violence: Creating a dialogue around vulnerable populations.* (pp. 157–241). New York: Nova Science Publishers.

Williams, D. J., Neville, F. G., House, K., & Donnelly, P. D. (2013). Association Between Old Firm Football Matches and Reported Domestic (Violence) Incidents in Strathclyde, Scotland. *SAGE Open*, *3*(3).

Woodward, K. (2017). Women's time? Time and temporality in women's football. *Sport in Society*, *20*(5-6), 689–700.

Wu, H., Wang, Y. J., & Cheng, X. (2008). Incremental Probabilistic Latent Semantic Analysis for Automatic Question Recommendation. In *Recsys'08: Proceedings of the 2008 acm conference on recommender systems* (pp. 99–106). Lausanne, Switzerland: ACM.

Yakubovich, A. R., Stöckl, H., Murray, J., Melendez-Torres, G. J., Steinert, J. I., Glavin, C. E., & Humphreys, D. K. (2018). Risk and Protective Factors for Intimate Partner Violence Against Women: Systematic Review and Meta-analyses of Prospective–Longitudinal Studies. *American Journal of Public Health*, *108*(7), e1–e11.

Yang, S., & Lynn, M. (2014). More Evidence Challenging the Robustness and Usefulness of the Attraction Effect. *Journal of Marketing Research*, *51*(4), 508–513.

Yu, X., Liu, Y., Huang, X., & An, A. (2012). Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain. *IEEE*

Transactions on Knowledge and Data Engineering, *24*(4), 720–734.

Zhen, S., & Yu, R. (2016). The development of the asymmetrically dominated decoy effect in young children. *Scientific Reports*, *6*(22678).