Author/s:
Pryor, Campbell Gregor

Title:
Towards understanding choice biases

Date:
2019

Persistent Link:
https://hdl.handle.net/11343/234047

UNIVERSITY OF MELBOURNE

DOCTORAL THESIS

Towards Understanding Choice Biases

Author:
Campbell PRYOR

Supervisors:
Assoc. Prof Piers HOWE
Assoc. Prof. Amy PERFORS

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Melbourne School of Psychological Sciences
Medicine, Dentistry & Health Sciences

January 28, 2020

Abstract

Melbourne School of Psychological Sciences
Medicine, Dentistry & Health Sciences
UNIVERSITY OF MELBOURNE

Doctor of Philosophy

Towards Understanding Choice Biases

by Campbell PRYOR

Our choices can often be biased in systematic ways. In this thesis, we study the cognitive mechanisms driving two prominent choice biases: the endowment effect and the descriptive norm effect. The endowment effect, a tendency to assign higher value to something you own, is typically explained based on loss aversion. A finding that the endowment effect reverses for choices between undesirable options has been taken as evidence against loss aversion. We find that these reversals are driven by people's expectations and only seem to occur when comparisons between the options are limited. We outline a model of loss aversion that can qualitatively and quantitatively account for these results. The descriptive norm effect, a bias towards popular options, is typically explained based on the fact that descriptive norms offer useful information as to which option is best. We show that descriptive norm effects occur even when the norm is entirely arbitrary and thus offers no useful information. Another prominent explanation of the descriptive norm effect, self-categorization theory, can explain the impact of these arbitrary norms. However, this theory predicts that people will avoid conforming to descriptive norms from an outgroup. We find this not to be the case, with participants' preferences shifting towards outgroup descriptive norms. We discuss how mechanisms from the broader choice bias literature could in principle explain our findings.

Declaration of Authorship

I, Campbell PRYOR, declare that this thesis titled, "Towards Understanding Choice Biases" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at the University of Melbourne.
- This thesis comprises only my original work, conducted towards completion of the degree of Doctor of Philosophy, except where indicated in the preface.
- Where I have consulted the published work of others, this is always clearly attributed.
- This thesis is fewer than 100,000 words, exclusive of figures, tables, footnotes, references and appendices

Signed:
Campbell Pryor
Date:
January 28, 2020

Preface

This thesis contains a compilation of original work, parts of which have been published as papers. The papers included in this thesis are detailed in the *Papers Included* section and have been adapted to enhance the readability of this work.

Paper 1 was published in *Judgement and Decision Making* in May 2018 and is adapted to form the contents of Chapter 4 of this thesis. Paper 2 was published in *Nature Human Behaviour* and is adapted to form the contents of Chapter 6 of this thesis. Paper 3 was published in *PLOS ONE* and is adapted to form the contents of Chapter 8 of this thesis.

For all of this work, testing, data collection, data analysis and written drafts were completed by the PhD Candidate. Drafts were revised and edited by the candidate and his supervisors, A/Prof. Piers Howe and A/Prof. Amy Perfors.

This research was supported by an Australian Government Research Training Program Scholarship.

Papers Included

Paper 1

Pryor, C. G., Perfors, A., & Howe, P. D. L. (2018). Reversing the endowment effect. *Judgement and Decision Making*, *13*(3), 275–286.

Paper 2

Pryor, C. G., Perfors, A., & Howe, P. D. L. (2019b). Even arbitrary norms influence moral decision-making. *Nature Human Behaviour*, *3*, 57–62.

Paper 3

Pryor, C. G., Perfors, A., & Howe, P. D. L. (2019a). Conformity to the descriptive norms of people with opposing political or social beliefs. *PLoS ONE*, *14*(7), e0219464.

Acknowledgements

I would like to begin by expressing my eternal gratitude to my supervisors, A/Prof. Piers Howe and A/Prof. Amy Perfors. Your enthusiasm for science is contagious and I aspire to critically engage with research at the level that you do. Your mentorship, encouragement and support has been beyond valuable to me. Thank you to my thesis committee, A/Prof. Stefan Bode and A/Prof. Daniel Little, for offering your wisdom and guidance at the times when I most needed it. I must also thank the incredible friends I have made throughout this process, Marcellin Martinie, Larson Landes, Jess Marris, Annie Blunden, Weijia Chen, Shi Xian Liew, Kathryn Hull, Sarah Moneer, Xue Jun Cheng, Dave Griffiths, Geoff Saw, Felix Singleton Thorn, Andrew Legg, Ariel Goh, Gabe Ong and the many other amazing scholars of Redmond Barry Building. I am truly lucky to have come across such a weird, wonderful and worryingly intelligent group of friends. To my superlative family, I would not be where I am without your support, encouragement and 'loving neglect'. Finally, to Natalie, thank you for giving me purpose, for pushing me to be more than I could ever be without you and for always ensuring I was at my happiest even when the PhD was at its hardest.

Contents

List of Figures

Chapter 1

Introduction

When presented with a choice with no clearly correct option, how do you arrive at a decision? For example, when deciding which of two meals to eat, what drives you to choose one over the other? In theory, this decision should be based on which meal will grant you the greatest utility, providing you with suitable levels of sustenance, satiation and satisfaction. In reality, such decisions can be influenced by a range of seemingly minor factors, such as what other people are eating (Burger et al., 2010), the other options available (Sen, 1998), or even simply the order in which you process the options (Biswas, Grewal, & Roggeveen, 2010; Bucher et al., 2016). The various ways in which our choices can be influenced by these secondary factors are referred to as choice biases and have fostered research interest across fields of psychology, behavioural economics, government policy and marketing.

There are a number of reasons these choice biases have gained substantial research interest. From a marketing or social intervention standpoint, these biases offer avenues in which people's choices can be influenced towards more desirable behaviour. People often make suboptimal decisions and so understanding how these choices are biased allows for choice contexts to be reframed such that they encourage more prosocial choices (Lehner, Mont, & Heiskanen, 2016). From an economic perspective, these biases can help us predict human behaviour, allowing for more accurate economic modelling and forecasting. For example, agent-based models can use what we know about human preferences and biases to simulate the potential effects of changes in economic conditions, which could otherwise not be inferred from historical trends or traditional economic models (Farmer & Foley, 2009).

From a psychology point-of-view, these biases offer great value as a tool for understanding how people make decisions. In a similar way to how visual illusions offer us insight into the visual system, choice biases offer us insight into the mechanisms driving choices. At a general level, the fact that choices can be influenced by context tells us that making a choice is not as simple as accessing some pre-existing value assigned to each option in order to determine which option is best. Instead, our preferences seem to often be constructed at the time of being presented with a choice (Slovic, 1995; Kahneman & Tversky, 1979). By better understanding the mechanisms underlying choice biases, we can better understand how people's choices are constructed and the resulting ways in which choices can be improved. Accordingly, this thesis focuses on the mechanisms underlying a number of choice biases.

As a background to studying these choice biases, we begin with a literature review (Chapter 2), looking more broadly at the cognitive mechanisms that have been proposed to explain various choice biases. The literature on different choice biases is largely disparate, and these biases are often treated as being highly distinct. However, they all relate to the same process of making choices and potentially overlap in many ways. Consequently, we consider it important to begin with a broad overview of this literature, given the potential for our understanding of specific choice biases to be informed by broader research on other choice biases. This review of the literature identifies a number of general mechanisms that could potentially explain multiple choice biases.

In terms of choice biases being treated as distinct, one trend that becomes apparent in the literature review is the strong tendency to explain each individual choice bias based on the salient, unique aspects of that bias. While this has allowed for a range of theories to be developed across the literature, it has also meant that there has been limited focus on what choice biases have in common.

In Chapter 3, we present a pilot exploration that compares multiple choice biases in terms of how they are affected by whether the choice stimuli are desirable (e.g. winning money) or undesirable (e.g. losing money). Studying undesirable choices is interesting because there is evidence that some choice biases may reverse when choosing between undesirable rather than desirable options (Armel, Beaumel, & Rangel, 2008; Brenner, Rottenstreich, Sood, & Bilgin, 2007; Houston, Sherman, & Baker, 1989). For example, Brenner et al. (2007) found that the endowment effect (a bias towards keeping an option that is already owned) reversed when choosing between undesirable options, wherein participants wanted to switch away from an undesirable option more when they were endowed with ownership of it than when they were not. These potential reversals are interesting because a large number of prominent theories do not predict such reversals to occur. Thus, by seeing whether various choice biases reverse for choices between undesirable options, we can better understand the mechanisms driving those biases. If certain choice biases do reverse and others do not, this could indicate that they are driven by meaningfully distinct cognitive processes.

Surprisingly, we not only failed to find a reversal between desirable and undesirable choice scenarios for any of the other studied choice biases, but we also failed to replicate the endowment reversal; our participants preferred keeping the endowed option, regardless of desirability. This led us to study endowment reversals in much more detail in Chapter 4. What we found was that the endowment reversal only occurred when comparisons between the available alternatives were limited. Furthermore, these reversals appear to be a function of participants' expectations, with reversals occurring even for desirable options if a participant was manipulated to have higher expectations.

While our results suggest that studying undesirable choices is a less powerful test of theories than expected, they also provide the important insight that one of the more prominent explanations of the endowment effect remains a valid theory. Furthermore, the fact that we did not find endowment reversals under more natural choice conditions argues against a couple of mechanisms that did predict choice biases to reverse for undesirable choices.

A second disparity in how various choice biases have been studied, is the extent to which research into different choice biases has focused on understanding the mechanisms underlying the bias. While the mechanisms underlying some biases (such as the endowment effect) have been extensively studied, other choice biases have received a far more applied research focus. Arguably the strongest example of this is the descriptive norm effect, the finding that we tend to show greater preference for an option when we know it is popular amongst other people (Asch, 1951; Sherif, 1936). Most of the descriptive norm effect research has focused on the real-world implications of people's tendency to follow norms. For example, studies have found that descriptive norms can be used to decrease tax evasion (Wenzel, 2005), increase organ donor registrations (Behavioural Insights Team, 2013) and decrease energy use (Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007). While such a focus is valuable in helping to further efforts to encourage prosocial decisions, it does mean that our theoretical understanding of the descriptive norm effect is underdeveloped. Chapter 6 and Chapter 8 attempt to address this limitation.

Potentially one of the reasons the descriptive norm effect has received little empirical theory-testing is that there are straight-forward, objective reasons for why people should show this effect. Two of the most prominent theories of the descriptive norm effect revolve around the idea that descriptive norms provide us with information that is relevant to our choices. The informational account proposes that whenever a decision maker has uncertainty about which option is best, they can turn to other people's decisions as a source of useful information. The popularity of a target option over a competitor implies that other people have determined that the target option is superior and the decision maker should therefore favour it too (Cialdini, Reno, & Kallgren, 1990; Deutsch & Gerard, 1955). The social sanction account offers another objective reason for why people conform to descriptive norms. If other people favour a particular option, they may respond negatively towards you (enact negative social sanctions) if you deviate from these preferences, offering an objective reasons why you should follow the descriptive norm, so as to avoid these sanctions (Deutsch & Gerard, 1955; Ajzen & Fishbein, 1980; Bendor & Swistak, 2001).

Chapter 6 looks at whether these prominent, objective accounts are required to explain the descriptive norm effect. By presenting norms that are said to have arisen for random reasons beyond anybody's control rather than reflecting people's actual preferences, we present norms that are entirely arbitrary and provide no objectively rational reason to follow them. Despite this, we repeatedly find that people continue to follow such norms, indicating that they are biased by descriptive norms even when there is no objective reason for such a bias. As a result, we conclude that alternative mechanisms are required to explain the descriptive norm effect.

Self-categorisation theory proposes one such alternative mechanism. Self-categorisation theory proposes that people's sense of identity is often driven by salient social categorisations (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). If people identify with a particular group (an ingroup) they may be driven to follow the norms of that group in an effort to maintain their personal sense of identity as a member of that group (Terry & Hogg, 1996). While self-categorisation can explain the effect of arbitrary norms studied in Chapter 6, it also predicts that people's tendency to follow these arbitrary norms will be influenced by whether they identify with the group that the norm relates to. Specifically, self-categorisation theory suggests that people will want to conform to ingroup norms and try to avoid conforming to outgroup norms (Hogg, Turner, & Davidson, 1990). This leads to the prediction that people will conform to ingroup norms more when also presented with an opposing norm that an outgroup tended to favour a different option.

Counter to this, in Chapter 8 we find that presenting an opposing outgroup norm alongside the ingroup norm actually shifts people's preferences towards the outgroup norm compared to when only the ingroup norm is presented. This result is inconsistent with self-categorisation theory. Thus, across Chapter 6 and Chapter 8, we demonstrate that none of the three most prominent accounts of the descriptive norm effect offer a full and complete picture of this effect. The prominent theories of the descriptive norm effect, and many other choice biases, focus heavily on the various salient, unique aspects of the effect. We discuss how more general mechanisms, which may apply across multiple choice biases, could offer a potentially more complete and parsimonious account of these biases, including the descriptive norm effect.

Bringing these results together, we conclude this thesis by looking at the broader implications these findings have for our understanding of how people make choices. This thesis emphasises the importance of understanding how people are processing the available information when testing theoretical explanations of choice biases (as evidenced by our study of the endowment effect reversal). It also exemplifies the importance of rigorously testing theoretical explanations of choice biases that might otherwise seem to have obvious explanations. Through these detailed studies of the endowment effect and the descriptive norm effect, this thesis offers new insights into the mechanisms driving people's choices.

1.1 Key definitions

Choice bias Choice biases are often defined as any case where preference between two or more options is inconsistent with normative models of how people should rationally behave (Tversky & Kahneman, 1974; Mukherjee & Srinivasan, 2013; Newell & Bröder, 2008). However, we will show in this thesis that, while a choice may seem consistent with rationality, it can actually be caused by biases that are inconsistent with objective rationality. We therefore adopt a broader definition of choice biases as any case where changing some aspect of a choice scenario can result in increased preference for a particular option. This definition has the advantage of being atheoretical, not making assumptions about whether rationality underlies a shift in preference.

Manipulation A useful way to think of choice biases is to imagine a baseline choice between a number of options. For example, "Here are two options with some information about their attributes. Which would you prefer?" A choice *manipulation* then involves changing that baseline choice in any way such that people tend to prefer a particular option more than they would have in the baseline choice or more than they would have if the manipulation instead favoured another option. For example, manipulating the baseline choice by informing the decision maker that a particular option has limited availability may increase preference for that option, a finding referred to as the scarcity bias (Lynn, 1989). In this case, the manipulation is the addition of information that one option has limited availability. We use this concept of manipulations as a useful way to define choice biases. In reality, choice biases can often occur naturally without anyone explicitly manipulating the choice but we can still imagine how an equivalent choice bias would be induced by manipulating a baseline choice.

Target In choice biases, the manipulation typically relates to a particular option. For example, the scarcity bias outlined in the above definition referred to a particular item as having limited availability. We refer to this option as the *target*. In some cases, a manipulation does not specifically relate to a particular option but nevertheless boosts preference for a particular option. In such cases, it is standard to refer to the option whose preference increases due to the manipulation as the *target*.

Competitor/alternative In contrast to the target, other options (whose preference share changes relative to the target) are referred to as *competitor* or *alternative* options.

Chapter 2

Literature Review: The Mechanisms Driving Choice Biases

A reasonable approach when faced with a difficult decision is to focus on the outcomes of each option and choose whichever option is likely to offer the greatest benefits. Indeed, for a long time, the standard approach to understanding choices was to offer normative theories of how people should act, identifying various axioms that people should follow in order to maximise their well-being (Arrow, 1951; Fishburn, 1970). Any reasonable person was then presumed to follow these axioms so as to maximise their wellbeing (von Neumann & Morgenstern, 1953). Termed utility theories (e.g. expected utility theory; von Neumann and Morgenstern, 1953), these approaches presumed that people have stable, well-defined preferences that can be represented by utility functions (Friedman & Savage, 1948; Pratt, 1964). An individual's pre-existing utility functions determine the relative value they derive from different options and it is assumed that the individual will choose whichever option maximises their expected utility. This standard assumption of neoclassical economics is foundational to widespread assumptions that have shaped economic systems: people should be willing to pay for something based on the value they would derive from that thing and people will exchange resources to achieve maximally-beneficial allocations of resources (Pareto optimality) whenever transaction costs are sufficiently low (Coase, 1960).

Normative theories can indicate how rational decision makers should act. However, there is substantial evidence that people do not behave in line with these theories. Axioms outlined by utility theorists dictate that people's choices should only be influenced by information that is relevant to the outcomes of their choice. Counter to this, it has repeatedly been found that people's choice can be biased by information that normative models would consider irrelevant (see McFadden, Machina, & Baron, 1999, for a review). Thus, these choice biases indicate that alternative, non-normative models are needed to explain people's choices.

This chapter begins with a review of evidence that people violate these normative models, justifying the need for alternative models of human decision making. Through then reviewing the wide range of non-normative theories that have been applied to explain different choice biases, we identify a number of recurring mechanisms, raising the question of whether certain biases might be explained by a common set of mechanisms (which we begin to test in Chapter 3 and Chapter 4). We also identify that certain choice biases have received far less empirical testing of these theories, leaving open the possibility that these choice biases could be explained by one of these identified mechanisms (which we test in Chapter 6 and Chapter 8).

2.1 Violations of normative models

McFadden et al. (1999) offer an excellent overview of many of the biases that have been observed in a range of judgement and decision making contexts. They provide a detailed commentary on why normative approaches to human decision making must adapt in light of these biases. To help justify why choice biases are so important to our understanding of human decision making, we now offer a somewhat briefer review of how biases represent evidence against normative models, with a particular focus on preferential choice biases.

Perhaps one of the strongest early cases against normative models was Tversky and Kahneman's (1974) exposé on how people's probability judgements fail to account for key statistical properties such as base rates and sample sizes. For example, participants were informed that 70% of people in a group were engineers and 30% were lawyers. However, when they were then presented with a description of a single member of the group, called Dick, that offered no indication of whether he was an engineer or a lawyer, participants indicated that Dick was just as likely to be a lawyer as an engineer, completely ignoring the base rate. Given integration of probabilistic information is core to normative models (Arrow, 1951; Fishburn, 1970), the distinct biases in people's probability estimates is problematic for utility theories. Violations of normative models have since been identified in a range of preferential choice scenarios.

An implicit assumption of utility theories is the idea of reference independence. Reference independence is the fairly straight forward assumption that people's preferences should be based on the final outcome of a decision rather than their state before the decision (Coase, 1960). In other words, if a decision results in the same outcome regardless of your starting point then your starting point should have no impact on your preferences. This is famously expressed by Coase theorem, which proposes that when people are able to freely trade resources, they will end up at an optimal final allocation of resources, regardless of the initial allocation of property ownership (Coase, 1960). In simpler terms, if I value a mug more than you do, then you receiving ownership of the mug should simply mean I buy the mug from you.

Kahneman, Knetsch, and Thaler (1990) found that these trades of mugs occurred far less than Coasian bargaining would lead to. Instead, it has repeatedly been found that people tend to want to keep options that they already own even if the costs of swapping to an equally-desirable alternative option are minimal (Knetsch, 1989;

Kahneman et al., 1990; Morewedge & Giblin, 2015), a finding referred to as the endowment effect (Thaler, 1980). In fact, it has been found that people's preferences between two options can reverse when they are arbitrarily endowed with ownership of one of the options (Knetsch, 1989; Shu & Peck, 2011). Knetsch (1989), for example, endowed participants with ownership of either a mug or a candy bar. When asked if they would like to switch their endowed item for the other, non-endowed item, 89% of participants endowed with a mug chose to keep the mug rather than switch it for the candy bar, while 90% of participants endowed with the candy bar chose to keep the candy bar over the mug. Similar reversals have been found for people preferring to stick with a default option (Johnson, Hershey, Meszaros, & Kunreuther, 1993) or, more generally, favouring status quo options over equally-desirable alternatives (Samuelson & Zeckhauser, 1988).

These results cannot be explained simply based on there being greater transaction costs for switching away from a status quo. This is evidenced by the fact that the studies cited above controlled for transaction costs. Meanwhile, other studies have shown that status quos can influence preferences between other options, even when the status quo itself is not chosen (Dhingra, Gorn, Kener, & Dana, 2012; Dean, Kıbrıs, & Masatlioglu, 2017; Herne, 1998). Instead, these findings represent a violation of reference independence; people are influenced by the status quo even if the outcome of the decision is the same regardless of which option is established as a status quo. This is inconsistent with utility theories and indicates that people's preferences are constructed at the time of making a choice.

Another rational axiom derived by utility theories is that people's preferences should be independent of irrelevant alternatives. If you would rather eat a salad than a burger, then knowing that a sandwich is also available should not suddenly result in you preferring the burger over the salad. In other words, a rational actor's preference between two options should not be impacted by the addition of a third option. Counter to this independence of irrelevant alternatives axiom though, a range of choice biases have been identified wherein adding a third alternative (termed a decoy) to a choice set can cause people's preferences between the original two options to reverse.

Decoy effect research has found that presenting a decoy option with certain attribute values relative to two choice options (see Figure 2.1 on the following page) can influence people's preference between these options. We refer to the option whose preference share tends to increase due to the addition of a decoy as the target and the other option as the competitor. Presenting a decoy that is similar to the competitor, for example, tends to increase the relative preference share of the target option over the competitor (Becker, DeGroot, & Marschak, 1963). This is not too surprising given that people who would prefer the competitor may simply now be split between choosing the competitor and the similar decoy, decreasing the relative choice share of the competitor compared to the target option.

More interesting are findings that people will tend to prefer the target option

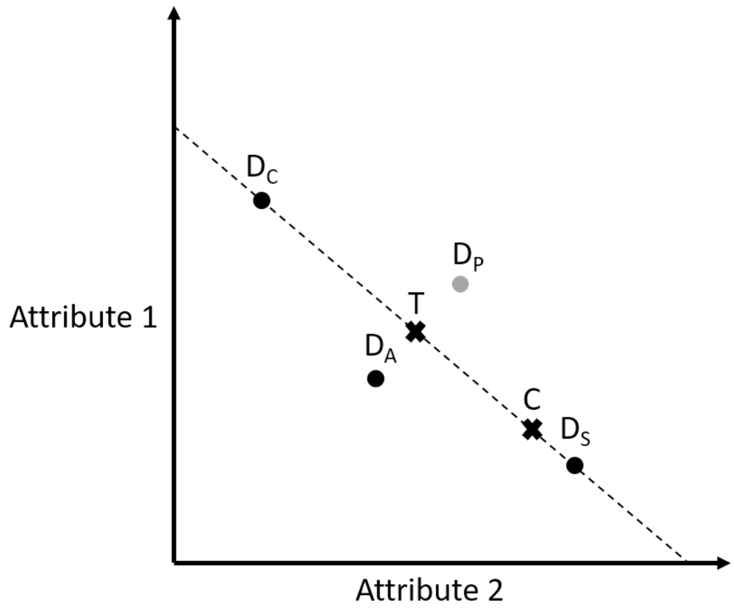

FIGURE 2.1: Examples of the relative attribute values used for different decoy effects, where higher values on either attribute are desirable. The crosses represent the target (T) and competitor (C) while the dots represent the potential decoys: attraction decoy (D_A), phantom decoy (D_P), similarity decoy (D_S) and compromise decoy (D_C).

when a decoy is offered that is similar to the target but inferior to it (termed the attraction effect; Huber, Payne, and Puto, 1982) or is similar to the target but unavailable for choosing (termed the phantom decoy effect; Pratkanis and Farquhar, 1992; Pettibone and Wedell, 2000). Despite these decoys not being chosen, their mere presence can boost preference for the similar target option. Another decoy effect, the compromise effect (Simonson, 1989), involves boosting preference for a target option by presenting an extreme decoy that positions the target option as a midpoint between the decoy and the competitor (see Figure 2.1). Given these decoys are not chosen and do not change the utility of the target or competitor options, utility theory cannot explain why people's preference are influenced by the presence of these decoy options. So, if people are not always making choices based on maximising their utility, how are they making choices? What mechanisms are driving these choice biases?

2.2 Bias-specific mechanisms

A trend that appears across the choice bias literature is the tendency to treat each choice bias as quite distinct, focusing heavily on the unique aspects of each different choice bias. This tendency is even seen in terms of how these biases are studied. For example, decoy effects such as the attraction, phantom decoy and compromise effects have tended to be studied using precisely described stimuli varying along two attributes (Huber et al., 1982; Pettibone & Wedell, 2000; Simonson, 1989). In contrast, the endowment effect is typically studied using relatively trivial but incentivised decisions relating to physically-present objects (Knetsch, 1989; Kahneman et al., 1990; Drouvelis & Sonnemans, 2017). Other choice biases are again studied differently to these biases, perhaps relying on preference between visual stimuli (e.g. the mere exposure effect discussed in Section 2.2.3 Zajonc, 1968; Johnson, Häubl, & Keinan, 2007; Newell & Shanks, 2007; Stafford & Grimes, 2012; Tom, Nelson, Srzentic, & King, 2007) or looking at applied contexts, using realistic but less controlled choice stimuli (Cialdini et al., 1990; Passafaro, Livi, & Kosic, 2019; Rimal & Real, 2003; Vinnell, Milfont, & McClure, 2019, e.g. social norm effects effect discussed in Section 2.2.4).

More interesting for our purposes though, this tendency to treat choice biases as distinct also manifests in the form of completely different theoretical explanations frequently being applied to different biases. There is a strong tendency in the literature for research to focus on the unique, salient aspects of a single choice bias when explaining it. For example, there has been a number of relatively recent papers arguing that the endowment effect is due to people inherently valuing ownership (Shu & Peck, 2011; Morewedge, Shu, Gilbert, & Wilson, 2009; Dommer & Swaminathan, 2013). While potentially true, such theories have the limitation of only applying to the endowment effect, not even being able to generalise to other status quo biases that do not involve ownership.

A major downside of this approach is that it results in an increasingly complex picture of human decision-making being built up, where different, but often similar, choice scenarios are assumed to engage completely different mechanisms. From the perspective of Occam's razor, being able to explain a wider range of biases under a smaller set of mechanisms is preferable. Without evidence that these biases are driven by distinct mechanisms, we should be naturally drawn towards testing whether more general mechanisms could potentially explain multiple choice biases.

We now review how a number of notable choice biases have been attributed to specialised, bias-specific mechanisms. Our attempt here is not to offer a comprehensive review of every theory offered for every choice bias. Instead, we hope to give some insights into prominent explanations of some major choice biases and exemplify a common tendency towards very bias-specific explanations. Later in this chapter, we look at how more generalisable mechanisms (e.g. loss aversion, attribute weighting etc.) could account for these and other choice biases.

2.2.1 Scarcity bias

The concept of scarcity (i.e. limited availability) is fundamental to micro-economic theory, wherein the cost of a good is driven up when demand outpaces supply (i.e. scarcity). In traditional economic theory, it is assumed that people's preferences are affected by scarcity only indirectly, to the extent that it influences the cost of goods. However, experimental studies have found that people are actually more likely to favour a good, simply when they know that its availability is limited, whether due to high demand, low supply, artificial restrictions, etc. (for an extensive review, see Oruc, 2015).

Prominent explanations of the scarcity bias focus heavily on the idea that people uniquely value scarcity, whether because we like to be unique (Brock, 1968; Fromkin, 1970) or believe that scarcity reflects value (Lynn, 1992, 1989). Support for these theories comes from correlations between scarcity and various measures of the assumed mediating variables. For example, Wu, Lu, Wu, and Fu (2012) and Chen and Sun (2014) found that stating an item was scarce led participants to offer higher ratings of perceived expensiveness, quality, value and uniqueness of the item. Lynn and Harris (1997) found that people with high need for uniqueness showed stronger scarcity biases. Williams, Saffer, McCulloch, and Krigolson (2016) found that scarcity tended to elicit neural responses associated with reward. Though consistent with the idea that scarcity is itself valued, the fact that the proposed mechanisms were not experimentally manipulated in these studies limits our ability to make causal conclusions. It remains possible that scarcity is engaging some alternative mechanism, which both increases preference for an option and, independently, increases perceived price, quality etc. of the scarce option.

If scarcity is implying higher value of an option, we should expect this to occur regardless of an individual's initial attitude towards the option. Counter to this, it has been found that people evaluated less desirable options more negatively when they were scarce (Ditto & Jemmott, 1989; Sehnert, Franks, Yap, & Higgins, 2014; Zhu & Ratner, 2015). Though these studies looked at ratings of scarce items rather than preferential choice, the fact that less desirable items are rated worse when they are scarce is inconsistent with the standard theories, wherein scarcity increases the perceived value of an option. Mechanisms that can account for this amplification are needed. In fact, the potential for manipulations to have an opposite effect on preference for an undesirable, rather than desirable option, has theoretical implications for a range of other choice biases, as studied further in Chapter 3 and Chapter 4

It has been suggested that a sense of resource unavailability due to scarcity heightens arousal (Sehnert et al., 2014; Zhu & Ratner, 2015). This is believed to increase engagement with the choice and thus, amplify the valence of all options. While this theory of general arousal can explain the amplification of ratings (more desirable items become more positive, less desirable options become more negative) when an option is scarce, the idea that scarcity induces a general sense of arousal means that one option being scarce will amplify the valence of all options, including the non-scarce options. Thus, this theory cannot explain the basic finding that people tend to favour a scarce option over an otherwise equally desirable non-scarce option. Instead, it may be that scarce options tend to attract more attention (Folger, 1992) which subsequently increases the valence of the scarce option. We discuss the idea that attentional processes might be driving a range of choice biases in Section 2.3.3.

2.2.2 Decoy effects

As discussed earlier, decoy effects (also commonly referred to as context effects) refer to findings that the addition of a decoy option can affect the relative preference share between two other options. There are some obvious similarities between different decoy effects. For example, the attraction and phantom decoy effects both involve presenting a decoy option that is similar to a target option but is not chosen, either because it is inferior or unavailable respectively (Huber et al., 1982; Pratkanis & Farquhar, 1992). Despite the similarity between these effects, it is common to see these (and other decoy effects) explained by distinct theories, exemplifying the tendency to focus on specific mechanisms that only explain a single choice bias. For example, similarity substitution theory proposes that the phantom decoy effect occurs because we accept the target option simply because it is the closest alternative to the highest value, desired option (the phantom decoy; Pettibone and Wedell, 2000). This theory is dependent on the decoy being superior and unavailable and thus, obviously does not extend to the attraction effect, where the decoy is inferior and available. Other common explanations of the phantom decoy effect also rely on the unavailability of the phantom decoy, such as reactance theory (Farquhar & Pratkanis, 1993; Brehm & Brehm, 2013) or commodity theory (Pratkanis & Farquhar, 1992; Brock, 1968), preventing them from explaining other decoy effects.

A number of models have arisen that focus on explaining multiple decoy effects at once. Multialternative decision field theory focuses on the idea of lateral inhibition, where the perceived value of an option is influenced by the relative value of similar options (Roe, Busemeyer, & Townsend, 2001). This lateral inhibition is thought to explain the attraction effect based on the inferiority of the decoy option causing "negative interference", which boosts preferences for the similar target option. The compromise and similarity effects can also be explained based on a somewhat temperamental interaction of lateral inhibition with other mechanisms, the details of which are not crucial here.

A competing model, the multiattribute linear ballistic accumulator model, assumes that we engage in pairwise comparisons between options, where we are more likely to compare similar options (Trueblood, Brown, & Heathcote, 2014a). In the case of the attraction effect, this means that we should compare the target and inferior decoy options more often, where the superiority of the target over the decoy results in higher perceived value of the target (leading to the attraction effect). This model also assumes that we favour more central attribute values, leading to the compromise effect.

Testing of these models has relied primarily on evaluating the fit of these models to data from multiple decoy effects at once (Trueblood et al., 2014a; Berkowitsch, Scheibehenne, and Rieskamp, 2014; potentially to a fault, as argued by Liew, Howe, and Little, 2016). Despite this, they place no emphasis on explaining other choice biases; the mechanisms adopted by these models are dependent on having three options with certain patterns of attribute values and thus, they have no bearing on choice biases that occur in two-alternative contexts (e.g. the endowment effect or scarcity bias). Inexplicably, they also fail to explain phantom decoy effects, instead predicting the exact opposite effect. The lateral inhibition proposed by multialternative decision field theory would suggest that a superior phantom decoy should inhibit preferences for the similar target option. Meanwhile, according to the multiattribute linear ballistic accumulator model, the tendency to compare similar alternatives should decrease preference for a target option when compared to a superior phantom decoy. Instead, the multiattribute linear ballistic accumulator model would have to include a separate mechanism that is engaged by phantom decoys in order to account for the phantom decoy effect (Trueblood et al., 2014a, p. 191). In Section 2.3.1 and Section 2.3.4, we discuss how generalisable models (based on loss aversion and attribute weighting mechanisms respectively) have been used to explain multiple decoy effects while also being able to account for other choice biases as well.

2.2.3 Mere exposure effect

The mere exposure effect refers to the finding that simply having been shown an option before (i.e. having been exposed to it) is sufficient to increase preferences for that exposed option (Zajonc, 1968). Typically, this exposure involves visually presenting perceptual stimuli such as faces (Harmon-Jones & Allen, 2001; Huang & Hsieh, 2013; Newell & Shanks, 2007), brand logos (Stafford & Grimes, 2012) or images of toys (Tom et al., 2007). Accordingly, theories of this effect focus on the role that visual exposure plays in processing fluency (the ease of processing information) and the resulting effect that fluency has on preferences. Where prominent theories disagree is primarily in how fluency leads to positive affect. Fluency may then reducing feelings of uncertainty (Zajonc, 1968; Winkielman, Schwarz, Fazendeiro, & Reber, 2003), may be inherently rewarding (Winkielman & Cacioppo, 2001; Xiang Fang, Surendra Singh, & Rohini Ahluwalia, 2007) or may be misattributed to liking of the exposed option (Bornstein & D'Agostino, 1994; Xiang Fang et al., 2007).

Regardless of the exact mechanisms, all of these theories propose that exposure to an item makes it easier to process which then increases liking of the exposed item. Counter to this, Armel et al. (2008) found that exposing people to an aversive food item resulted in decreased preference for this exposed food relative to a non-exposed, equivalently aversive food. This finding suggests that exposure does not necessarily increase liking of an item but instead amplifies its valence; desirable items are more desirable but undesirable items are more undesirable. We discuss some more general mechanisms that predict this amplification in Section 2.3.3 and Section 2.3.4.

2.2.4 Social norm effects

Social norm effects are a category of choice biases that focus on how other people can influence our preferences. It is common to distinguish between descriptive norms and injunctive norms (Cialdini et al., 1990; Lapinski & Rimal, 2005).

The descriptive norm effect refers to findings that we will tend to favour an option more when we know that that option is popular amongst other people. In applied experiments, informing people of the relevant descriptive norm has been found to increase organ donor registrations (Behavioural Insights Team, 2013), decrease tax evasion (Wenzel, 2005), increase towel reuse at hotels (Goldstein, Cialdini, & Griskevicius, 2008), decrease long-term energy use (Schultz et al., 2007) and increase recycling behaviour (Czajkowski, Zagórska, & Hanley, 2019, e101110). For example, the Behavioural Insights Team (2013) found that informing drivers that thousands of other people sign up as organ donors each day significantly increased organ donor registrations compared to a baseline message just asking drivers to "please join the NHS Organ Donor Register."

Whereas descriptive norms involve what other people actually do, injunctive norms involve what people believe you should do (Cialdini et al., 1990). These injunctive norm effects can occur in a negative sense, where a choice is discouraged, such as Cialdini et al. (1990) finding that people littered less when presented with an injunctive norm asking people not to litter. They can also occur in a positive sense, where people will be more likely to choose an option when it is generally endorsed by others (Gupta & Harris, 2010; Senecal & Nantel, 2004; Shi & Wang, 2008).

Thus far, the choice biases we have discussed all represent violations of normative models, wherein decision makers are influenced by information that is irrelevant to the outcomes of their decision. Accordingly, given those biases cannot be accounted for by normative models, much of the research into these choice biases has focused on testing various mechanisms that might be underlying them. In contrast, empirical tests of the mechanisms underlying these social norm effects is rare. Instead, an extensive literature has developed looking at applying these biases in various contexts or looking at factors that moderate the strength of these effects. Just this year, we already have research looking at the impact of descriptive or injunctive norms on legislation surrounding natural disasters (Vinnell et al., 2019), cannabis use (Ecker, Dean, Buckner, & Foster, 2019), recycling (Passafaro et al., 2019) and tax compliance (Alm, Schulze, Von Bose, & Yan, 2019), to name a few.

It is quite likely that the heavy focus on applications of these biases and the minimal focus on studying the cognitive mechanisms causing them is at least partially due to the fact that there are a number of intuitive, objective reasons for people to conform to these norms. If we can objectively explain why someone should want to conform towards a norm then such conformity tells us little about that person's decision making process beyond indicating they are capable of the relevant, objectively rational thought processes. There are two prominent explanations of social norm effects that offer objective reasons why people should want to conform to norms.

The social sanction account focuses on the idea that violating a social norm is likely to result in people responding negatively towards you (Ajzen & Fishbein, 1980; Bendor & Swistak, 2001). Thus, in an effort to avoid aversive social sanctions, it is reasonable for a decision maker to conform with social norms unless an alternative option is clearly superior. Though this is most directly associated with injunctive norms, where there is a clear sense that others endorse a certain action, people tend to infer injunctive norms from the presence of descriptive norms (Eriksson, Strimling, & Coultas, 2015). Thus, it may be that norms generally imply social sanctions, changing the relative utility of different actions in favour of the social norm.

More closely associated with the descriptive norm effect is the informational account. This theory proposes that we conform to descriptive norms because they offer relevant information as to which option is best (Cialdini et al., 1990; Deutsch & Gerard, 1955). If most other people prefer a particular option, it is likely because they have deemed that that option is preferable for some reason. Thus, the choices of others is indicative of "effective and adaptive action" (Cialdini et al., 1990, p. 1015).

An alternative, more subjective explanation for these effects relates to people's sense of self-identity. Self-categorisation theory (and the related, social identity theory) propose that an individual's sense of identity can be heavily influenced by the social groups they identify with (Tajfel, 1970; Terry & Hogg, 1996). In order to maintain this sense of group identity, self-categorisation theory proposes that people will often want to engage in behaviours that "optimally minimise ingroup differences and maximise intergroup differences" (Terry & Hogg, 1996, p. 779), resulting in conformity to ingroup norms so as to minimise ingroup differences (Hogg et al., 1990).

Whether social norms impact preferences through informational, social sanction or self-categorisation mechanisms, we can see that all of these theories are heavily reliant on unique information offered by social norms. However, if we look at other choice biases, we can begin to recognise similarities between social norm effects and other choice biases. Accordingly, in Section 2.3 we discuss some more general mechanisms from research into other choice biases that could potentially explain these social norm effects.

It is worth noting that research interest into the mechanisms underlying social norm effects is not completely absent. A recent meta-analysis by Melnyk, Van Herpen, Jak, and Van Trijp (2019) hoped to get a sense of whether descriptive and injunctive norms affected preferences through different processes by looking at whether these effects were mediated by attitude, perceived behavioural control and behavioural intentions. They found that descriptive norms had a significant direct effect on behaviour even when controlling for these proposed mediators. They interpreted this direct effect as evidence that the useful information the descriptive norms provide has resulted in them being used as heuristics to help easily choose the best option. On the other hand, injunctive norms had a significant indirect effect but no significant direct effect, potentially supporting the theory that injunctive norms cause people to think more about the decision, which then indirectly affects behaviour via changing people's attitudes and intentions. However, the lack of a direct effect of injunctive norms is more likely a result of the weak, non-significant total effect of injunctive norms, rather than indicative of the mechanisms driving this effect. Thus, these results tell us less about the mechanisms underlying potential injunctive norm effects and instead raises questions as to the generalisability of the injunctive norm effect.

Another set of studies have looked at fitting computational models to social norm effects on visual decision making (Germar, Schlemmer, Krug, Voss, & Mojzisch, 2014; Germar, Albrecht, Voss, & Mojzisch, 2016; Germar & Mojzisch, 2019). We discuss these studies more in Section 2.3.5. As a brief summary, Germar and colleagues were interested in distinguishing whether social norms bias how people process visual information (a perceptual bias) or instead bias people to support a norm somewhat independent of the available information (a judgement bias). By including parameters in the model that theoretically represent these two separate processes, they found that the parameter representing a perceptual bias better accounted for the impact of social norms on visual judgements.

As a complement to these approaches, Chapter 6 and Chapter 8 of this thesis present experimental tests of some strong predictions made by the informational, social sanction and self-categorisation theories of the descriptive norm effect. Prior to the current thesis, we had little idea of the extent to which of the informational, social sanction or self-categorisation best explain social norm effects, if any. The studies of social influence discussed above did not explicitly test predictions of any of these three prominent mechanisms. We directly test these theories in Chapter 6 and Chapter 8.

In any case, for each of these prominent theories we again see a tendency to focus on the salient, unique aspects of these choice biases when explaining them. All of these theories rely entirely on the idea that social norms carry social information that directly influences preferences (whether through offering useful information, implying sanctions or reflecting group identity). We now discuss a number of more general mechanisms that could be underlying multiple choice biases.

2.3 Generalisable mechanisms

The theories discussed above all relied on unique aspects of the specific bias they attempted to explain. The specificity of these theories implies that the given choice bias they are trying to explain is unique and is driven by distinct mechanisms from other choice biases. This raises the question of whether such specific mechanisms are required. It seems plausible that a general set of mechanisms might act across a range of choice biases. Looking at the similarities between certain choice biases, it is easy to imagine how they might be driven by the same mechanisms. For example, though social norm effects are generally attributed to the unique implications of a social norm (such as the potential for social sanctions), they could also simply represent a status quo. Like endowments and defaults, social norms can represent a status quo of the currently expected behaviour. Thus, it could be that people conform to social norms for the same reasons they conform to status quo biases such as the endowment effect and default bias. Relying on bias-specific theories for each different choice bias builds up a complex picture of preferential choice, where certain mechanisms are engaged for a certain choice bias while completely different mechanisms are engaged for a different choice bias. We now discuss a number of more general mechanisms that offer potentially more parsimonious accounts of multiple choice biases. We return to these generalisable mechanisms throughout the thesis, especially in Chapter 9, where we consider the extent to which each of these generalisable mechanisms could explain our findings.

2.3.1 Loss aversion

Perhaps one of the most famous and widely applied theoretical mechanisms stemming from research into people's preferences is the notion of loss aversion. Loss aversion refers to the idea that experiencing a decrease in value (a loss) is considered to be more painful than an equivalent gain in value is considered pleasurable (see Figure 2.2 on the facing page). This idea was popularised by prospect theory, which proposed that people do not simply make decisions based on the absolute value of options but instead based on their value relative to a reference point (Kahneman & Tversky, 1979; Tversky & Kahneman, 1991). Because losses relative to this reference point are believed to loom larger than equivalently sized gains, this leads to the prediction that people will tend to favour options that are closer to their reference point. Various efforts to estimate the increase weight people place on losses than gains indicate that losses loom approximately twice as large as equivalently sized gains (Gächter, Johnson, & Herrmann, 2007; Tversky & Kahneman, 1992; Booij, van Praag, & van de Kuilen, 2010; Abdellaoui, Bleichrodt, & Paraschiv, 2007; Abdellaoui, Bleichrodt, l'Haridon, & Van Dolder, 2016), though loss aversion estimates have ranged from 1.43 (Schmidt & Traub, 2002) to 4.8 (Fishburn & Kochenberger, 1979).

A nice exemplification of how loss aversion can explain choice biases is seen in explaining the finding that people prefer options with small gains and losses over options that offer large gains and losses. Specifically, if there are two outcomes, x and y, such that $0 < x < y$, people will tend to favour a gamble with a 50% chance of gaining x and 50% chance losing x over a gamble with equal chances of gaining or losing y (Kahneman & Tversky, 1979). Even though both gambles have the same expected value ($0.5x - 0.5x = 0.5y - 0.5y = 0$), the larger potential losses of the latter gamble (given y is greater than x) seem to outweigh its larger potential gains. This preference for an option with small advantages and disadvantages over an option with large advantages and disadvantages has also been shown for non-risky choice stimuli, such as jobs (Tversky & Kahneman, 1991). From a loss

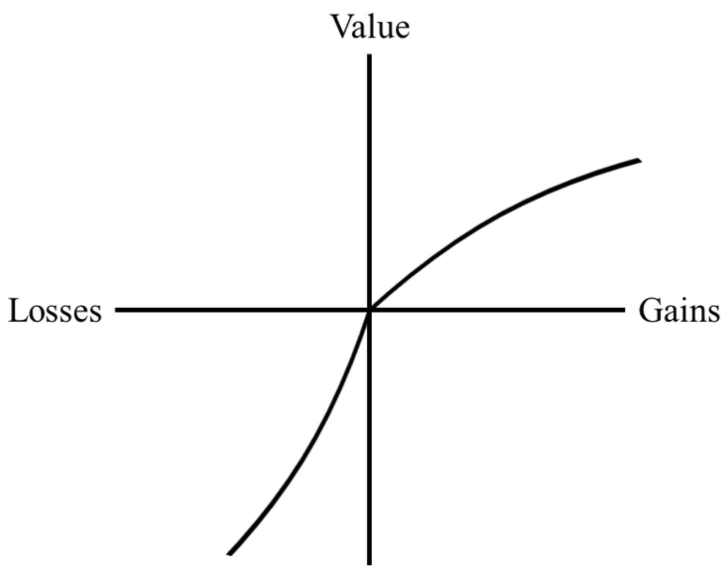

FIGURE 2.2: Visualisation of prospect theory's value function adapted from Kahneman and Tversky (1979). The x-axis represents objective gains and losses relative to a reference point (the origin) while the y-axis represents the perceived value of the corresponding gains and losses. Loss aversion is represented by the steeper slope in the losses domain than the gains domain.

aversion perspective, even though experiencing a large gain is pleasant, experiencing an equivalently sized loss is more painful than the gain is pleasurable. Thus, if two options offer gains and losses relative to a reference point (e.g. your current state) and are otherwise similarly valuable, loss aversion predicts that people will prefer the option that has smaller losses relative to their reference point.

Perhaps the choice bias most closely associated with loss aversion is the endowment effect (we value what we own), to such an extent that loss aversion is often treated as a description of the endowment effect (Novemsky & Kahneman, 2005). The endowment effect is essentially treated as an extreme example of the bias discussed in the above paragraph, wherein the endowment represents people's reference state, such that it has no gains and losses (Tversky & Kahneman, 1991; Kahneman & Tversky, 1979). Instead, all other option are evaluated in terms of their gains and losses relative to the endowed option. Consequently, an otherwise similarly desirable, non-endowed alternative will have some gains and some losses in value relative to the endowment. Loss aversion then suggests that the losses in value will have a bigger impact on preferences than the gains in value, such that the individual will not want to switch away from the endowed option.

A key advantage of loss aversion is its wide applicability. Beyond the endowment effect, loss aversion has been applied to a range of other findings, including multiple choice biases. Status quo biases generally, such as the default bias (the finding that people prefer to stick with a default) could be explained in much the same way as the endowment effect, wherein any status quo is set as a reference point where losses relative to the status quo are avoided. Though we are not aware of research applying loss aversion as an explanation for the scarcity, descriptive norm or injunctive norm effects, it is easy to imagine how loss aversion could apply to these effects in the same way. A scarce option is likely to run out if we don't choose it, so it may make sense to treat it as a reference point, evaluating whether other options offer sufficient gains to give up on obtaining the scarce option. Social norms establish a status quo of normal behaviour and so may similarly be treated as a reference point against which we evaluate alternatives, becoming averse to losing value relative to others (colloquially, a "fear of missing out").

Loss aversion has also been offered as a potential explanation of the attraction, compromise and phantom decoy effects, though in a slightly different conceptualisation to that proposed by prospect theory. Whereas prospect theory assumes that we have a certain reference point against which all other options are evaluated, the leaky competing accumulator model proposes that at any point in time we will attend to a particular attribute and look at the pairwise gains and losses all items have relative to each other on that attribute (Usher & McClelland, 2001; Usher & McClelland, 2004). As can be seen in Figure 2.1 on page 10, the attraction, phantom decoy and compromise effect all involve adding a decoy that is more similar to a target option than a competitor option. Thus, the competitor option has larger gains and losses relative to these decoys than the target option does. Due to loss aversion, these large relative losses should decrease preference for the competitor, explaining why attraction, phantom decoy and compromise decoys tend to boost preference for the nearer-by target option (Usher & McClelland, 2004; Tsetsos, Gao, McClelland, & Usher, 2012).

Neuropsychological evidence has supported the proposed impact of loss aversion in preferences. The amygdala has been linked with anticipation of losses (Kahn et al., 2002). Accordingly, De Martino, Camerer, and Adolphs (2010) found that patients with amygdala damage showed markedly less evidence of loss aversion when choosing between gambles that varied in their potential losses. Similarly, Sokol-Hessner, Camerer, and Phelps (2012) found that instances where healthy individuals showed greater aversion to options involving losses showed higher amygdala activation. Furthermore, ingestion of acetaminophen, a drug that inhibits negative sensations like pain and loss was found to weaken the endowment effect (DeWall, Chester, & White, 2015). These results all indicate that stronger neural responses to losses may play an important role in people's choices.

While loss aversion is a dominant theory in much of the decision making and judgement literature, Gal and Rucker (2018) offer an important reminder that it is just that, a theory. Gal and Rucker (2018) point out that there is a tendency to treat choice biases such as the endowment effect as evidence of loss aversion when instead we should be searching for evidence that loss aversion explains the endowment effect. Indeed, when going through evidence for loss aversion, Gal and Rucker (2018) show that much of the evidence is indirect. When direct measures have been attempted, the results have often been mixed. While Simonson and Kivetz (2018) temper many of the strong claims against loss aversion made by Gal and Rucker (2018), it remains quite possible that the biases often considered synonymous with loss aversion could indeed be attributed to other mechanisms.

One set of findings that loss aversion is weaker at explaining are findings that many choice biases occur even when people are choosing between options that do not involve clear consequences and thus, cannot result in any real losses to the participant. Various choice biases have been shown to occur even for visual stimuli that do not have a hedonic component (e.g. Fleming, Thomas, & Dolan, 2010; Trueblood, Brown, Heathcote, & Busemeyer, 2013; Trueblood & Pettibone, 2015; Trueblood, 2015). For example, Trueblood (2015) had participants try to choose the largest of two rectangles that varied in height and width. They found that designating one of these rectangles as a status quo choice led participants to judge it as larger. This was taken as evidence that status quo biases, such as the endowment effect, occur in the absence of loss aversion.

However, losses may not be limited strictly to cases where there are real losses in utility. People may experience a sense of loss whenever a change would hinder their ability to meet their aims during a choice, however abstract those aims are. If your aim is to pick the largest rectangle out of a set, then a rectangle being thinner than your reference point represents a loss in relation to that aim. This is relevant to hypothetical decisions, because if we limited our definition of losses to those with an impact on people's utility, loss aversion would be unable to explain any biases during hypothetical choices. In contrast, some of the earliest work on loss aversion relied on hypothetical scenarios. Galanter and Pliner (1974) had participants adjust an auditory tone based on how much more or less happy they would hypothetically be to receive a payout compared to a reference payout. Despite knowing that their responses would have no impact on their wellbeing, participants showed evidence of loss aversion, dropping off in happiness much faster for hypothetical losses than they increased for hypothetical gains. Consistent with this, Miyapuram, Tobler, Gregorios-Pippas, and Schultz (2012) found that participants showed similar neural responses to imagined, hypothetical outcomes as real outcomes, with these neural responses scaling based on the magnitude of the hypothetical outcome.

More concrete evidence against loss aversion was provided by Brenner et al. (2007), who found that the endowment effect reversed for choices between undesirable options. For example, they had participants imagine receiving a speeding ticket and either endowed them with a fine or having to attend traffic school. When endowed with the fine, participants tended to want to swap the fine to instead attend traffic school more so than participants endowed with attending traffic school. In other words, the endowment effect reversed, with participants wanting to swap away from their undesirable endowed option.

From the perspective of loss aversion, the absolute value of the options (whether they are desirable or undesirable) should not impact the endowment effect. Instead it is the value of the alternative relative to the endowment that matters according to loss aversion. A loss in value from a good attribute to a less good attribute should be equivalent to a loss in value from a negative attribute to a more negative attribute from the perspective of loss aversion. Consequently, this endowment reversal has been taken as evidence against loss aversion as an explanation of the endowment effect (Brenner et al., 2007).

Instead, Brenner et al. (2007) propose an alternative theory, they called possession loss aversion, as opposed to the standard "valence loss aversion" discussed above. This theory is perhaps better termed possession loss *amplification* because what it proposes is that the impact of losing possession of an item is exaggerated. Losing possession of a a desirable endowment is particularly bad while losing possession of an undesirable endowment is particularly good. Though possession loss amplification can explain Brenner et al.'s (2007) findings, it loses one of the strengths of loss aversion, which was its potential to explain a wide range of different choice biases. Other status quo biases, decoy effects, social norm effects etc. do not involve losing possession of an item and thus, cannot be attributed to possession loss aversion. We again see an example of the tendency towards bias-specific theories that focus on explaining an individual choice bias. In Section 2.3.3 and Section 2.3.4, we discuss two more general mechanisms (item attention and attribute weighting) that can potentially explain multiple choice biases while also predicting that these biases will reverse for choices between undesirable options. We study these reversals of choice biases for undesirable choices in Chapter 3 and Chapter 4, offering new insights into their theoretical implications.

2.3.2 Regret aversion

Similar to the idea of loss aversion is regret aversion. Regret aversion involves the fairly benign claim that people prefer not to choose an option that they will later regret (Kahneman & Miller, 1986; Baron & Ritov, 1994). Where regret aversion gets interesting is in the idea that various changes to how a scenario is presented can cause us to experience greater regret. More generally, Kahneman and Miller (1986, p. 145) propose that "the affective response to an event is enhanced if its causes are abnormal". For example, they found that participants expected someone to be less upset (show less regret) about crashing their car on their regular route home than on a new route that they only took for a change of scenery. In the context of choice biases, this notion that abnormal events elicit greater emotional reactions is commonly applied to explain the default bias.

People have a strong bias to stick with default options rather than switch to non-default alternatives (Johnson et al., 1993; de Haan & Linde, 2018; Shevchenko, Von Helversen, & Scheibehenne, 2014; Baron & Ritov, 1994; Ritov & Baron, 1990; Schweitzer, 1994). Baron and Ritov (1994) proposed that the default bias was at least partially due to people taking on greater responsibility for acts than omissions to act. In accord with Kahneman and Miller (1986), this greater responsibility amplifies the affective impact of both positive consequences and negative consequences resulting from an act, compared to an omission.

Similar to the notion of loss aversion, Ritov and Baron (1990) suggest that people's greater aversion to experiencing regret drives a preference for omissions over actions. Ritov and Baron (1992) supported the importance people place on acts vs omission by showing that people tend to prefer to switch away from a status quo when keeping the status quo requires an act. Consistent with the idea that people show greater regret from acts, Baron and Ritov (1994) found that acts which led to the worse of two possible outcomes were judged as worse than omissions that led to the exact same outcome. Nicolle, Fleming, Bach, Driver, and Dolan (2011) conceptually replicated this result, finding that experienced regret was greater when a sub-optimal outcome resulted from rejecting a status quo than from keeping the status quo. Additionally, they found that neural responses to this regret were associated with a stronger status quo bias in a subsequent choice.

Regret aversion has been applied quite specifically to status quo biases for risky prospects (i.e. gambles). However, it could generalise to other choice biases as well as non-risky choices. In relation to generalising to non-risky options, any case where a decision maker is uncertain as to the value they will derive from an option opens up the possibility for regret; deriving a lower value than expected may elicit regret. In relation to generalising to choice biases other than status quo effects, it is reasonable to assume that various other manipulations could establish an option as a reference point, where deviating from that option is treated as an act that has the potential to result in greater regret. For example, deviating from a descriptive norm requires going against how most other people act. This potentially results in taking greater responsibility for your actions when deviating from a norm, amplifying the potential for regret.

In Chapter 9 and Chapter 8, we discuss how regret aversion could potentially explain the findings from a range of the experiments presented in this thesis. Like loss aversion though, regret aversion cannot explain findings that some choice biases may reverse when choosing between undesirable (rather than desirable) options (Brenner et al., 2007; Armel et al., 2008). If anything, regret aversion would predict that choice biases should get stronger when choosing between undesirable options, given more negative outcomes tend to elicit greater regret than less positive outcomes (Baron & Ritov, 1994).

2.3.3 Item attention

Another general mechanism that could explain multiple choice biases relates to the amount of attention given to different options. It may be that various manipulations attract attention towards a target item and increased attention then affects preference for that item. An eye-tracking study by Shimojo, Simion, Shimojo, and Scheier (2003) found that when asked which of two faces they preferred, participants tended to look more at the face they ultimately ended up choosing. Shimojo et al. (2003)

proposed that looking at an option tends to increase preference for that option. This idea was formalised into a computational model by Krajbich, Rangel and colleagues 2010, 2011, 2012, termed the attentional drift diffusion model. This model proposed that people accumulate evidence for or against an item when looking at that item until eventually reaching sufficiently strong evidence in favour of one item that it is chosen. As with Shimojo et al. (2003), Krajbich et al. found that participants were more likely to choose an item the more it was looked at (relative to an alternative item).

Of course, it is possible that people simply tend to look at items that they prefer more, rather than preferring items that they look at more (i.e. causality may act in the opposite direction). As stronger evidence that attention to an item does indeed influence preference for that item, Shimojo et al. (2003) found that experimentally manipulating participants to gaze more at an item increased preference for that item. Furthermore, Simion and Shimojo (2007) found that increased time spent looking towards a chosen item did not continue after participants reported their decision. This indicates that people do not simply tend to prefer looking at items that they like (which should presumably continue even when a decision is not being made).

Armel et al. (2008) extended these finding to choices between undesirable options. According to item attention theories such as the attentional drift diffusion model, looking at a desirable option will accumulate evidence in favour of that option such that it is more likely to be chosen. In contrast, looking at an undesirable option will accumulate evidence against that option, such that it is less likely to be chosen. Consistent with this, Armel et al. (2008) found that exposing participants to an image of an aversive food for longer decreased the preference share of this option over another aversive food that was presented more briefly. This is consistent with the idea that increased attention towards an option polarises attitude towards that item (positively if the item is desirable, negatively if it is undesirable). In the same way, item attention theories can potentially explain Brenner et al.'s (2007) finding that the endowment effect reversed for choices between undesirable options. The question then becomes whether choice bias manipulations (such as endowing an option) actually increase attention towards the target option.

Consistent with item attention theories, some choice biases have been found to involve greater attention towards the target item. Geng (2016), for example, ran a Mouselab study of the default bias, wherein information about each option was hidden behind separate boxes and was only revealed while the participant hovered their mouse over each box. This allowed him to track how much time participants spent viewing the default option relative to non-default options. Geng found that setting an option as a default led to increased viewing time (36 seconds on average) compared to non-default options (24 seconds on average). The effect of the default was no longer significant once controlling for viewing time, indicating that time spent viewing may have accounted for the default bias. Dean et al. (2017) argued that item attention also drives the endowment effect, predicting that the endowment effect would be stronger when participants had greater attentional load (such that increased attention towards an endowed option is more important). Consistent with this, they found that increasing the number of available options (and thus the number of items that had to be attended) increased the strength of the endowment effect.

One potential limitation of these studies is that the experimental design encouraged increased attention towards the target item in a way that might not happen naturally. Dean et al. (2017) presented participants with the status quo option before any other options and subsequently placed it at the top of the screen. Thus, participants were exposed more to the endowed option, making it unclear whether endowments would naturally attract this attention. The use of Mouselab software by Geng (2016) may have also influenced how their participants attended to the available options. Restricting participants to only view one options at a time may have forced participants to evaluate the options separately. This may be different to how people would naturally process the options. For example, in the case of the attraction, compromise and similarity decoy effects, an eye-tracking study by Noguchi and Stewart (2014) found that participants' gaze tended to transition between items along the same attribute rather than between attributes on an individual item. This indicated that participants were naturally comparing items along one attribute at a time, rather than evaluating each item separately before deciding which to choose. Thus, it could be the case that attention to an option does tend to influence preference for that option but it is unclear whether such attention happens under natural conditions. Instead, Noguchi and Stewart's (2014) findings are more consistent with the idea that attention towards certain attributes rather than certain items could drive choice biases. We discuss this possibility in the next section.

2.3.4 Attribute weighting

In any choice situation, we can break down the available alternatives into a set of attributes that they differ on, either quantitatively (e.g. one car has better mileage than another) or qualitatively (e.g. one car is blue, the other is red). If you care more about one attribute over another, you will tend to favour the option that scores highly on your preferred attribute. While you may have some stable attitudes towards different attributes, attribute weighting mechanisms propose that various choice manipulations can change how you weight the available attributes, potentially explaining various choice biases (Ariely & Wallsten, 1995; Bhatia, 2013; Nayakankuppam & Mishra, 2005; Pachur & Scheibehenne, 2012; Bordalo, Gennaioli, & Shleifer, 2012a; Carmon & Ariely, 2000; Dhingra et al., 2012; Houston & Sherman, 1995).

For example, cancellation and focus theory is a popular explanation of the serial-order effect and relies on attribute weighting (Houston & Sherman, 1995; Sütterlin, Brunner, & Opwis, 2008). The serial-order effect refers to the finding that the order in which a decision maker evaluates the available options will influence their preferences (Houston et al., 1989; Biswas et al., 2010; de Bruin & Keren, 2003; Li & Epley, 2009). Houston et al. (1989) presented participants with a list of traits an option possessed, some positive and some negative (e.g. a person was described as boastful, lazy, cheerful and enthusiastic). They then presented participants with a list of traits about a second option. This second option either matched the first option on its negative traits (boastful and lazy) while having unique positive traits (e.g. attentive and considerate) or matched on the positive traits (cheerful and enthusiastic) while having unique negative traits (e.g. jealous and sloppy). They found that when the options differed on negative traits, participants tended to prefer the option presented first whereas when the options differed on positive traits, participants favoured the second option.

Cancellation and focus theory proposes that this occurs because all features are given similar weight when evaluating the first option (Houston & Sherman, 1995). In contrast, when evaluating the second option, the matching features are ignored, resulting in the unique features of the second option standing out and being given greater weight. Having unique positive features being weighted heavily leads to preference for the second option while having the unique negative features of the second option stand out decreases its perceived value.

Attribute weighting has been used to explain other choice biases. For example, Bordalo et al.'s (2012a) salience theory proposes that the endowment effect occurs because people tend to evaluate their endowment before non-endowed alternatives. When first evaluating the endowment, a decision maker tends to have little context for the choice. As a result, the attributes that the endowment scores highly on stand out. When going on to evaluate other options, the decision maker now has some context of what to expect, resulting in the attributes that the other option score highly not standing out. Thus, salience theory proposes that we tend to assign greater weight to the attributes that the endowed option scores highly on, leading to preference for the endowment when it scores highly on desirable attributes.

Bhatia's (2013) associative accumulation model extends this idea to explain multiple choice biases. Like salience theory, the associative accumulation model assumes that the target option in various status quo biases is salient, increasing weighting of the attributes that the target scores highly on. In the case of decoy effects, the attraction, phantom and compromise decoys all involve adding an item that scores highly on the same attribute as the target option. The greater number of items scoring high on that attribute is believed to increase weighting of that attribute, leading to increased preference for the target item (provided this attribute is positive).

From the perspective of salience theory and the associative accumulation model, any manipulation that makes a target item more salient will tend to result in the attributes that that item scores highly on standing out more and thus, being weighted more. It seems reasonable to assume that pre-exposing someone to an item (mere exposure effect) or informing them that an option is scarce (scarcity bias), popular (descriptive norm effect), recommended (injunctive norm effect), endowed (endowment effect) etc. will increase the salience of that item. Thus, the assumption that we assign greater weight to the attributes that stand out on salient items could potentially explain all of these choice biases.

An advantage of attribute weighting explanations is that they naturally explain findings that certain choice biases reverse when the choice is between undesirable rather than desirable options (Brenner et al., 2007; Armel et al., 2008; Houston et al., 1989). As discussed above, serial order effects reverse when items differ on negative attributes rather than positive ones (Houston et al., 1989; Biswas et al., 2010; Li & Epley, 2009). Brenner et al. (2007) found a reversal of the endowment effect when participants chose between undesirable options, as reviewed in Section 2.3.1. Armel et al. (2008) found that the mere exposure effect also reversed for undesirable stimuli, as reviewed in Section 2.2.3. Assigning greater weight to attributes that a target item scores highly on will lead to decreased preference for that item if it has high scores on negative, undesirable attributes. Thus, attribute weighting can explain these reversals.

Another strength of attribute weighting models is that they make quite clear predictions about how people attend to available information, which can be tested using process-tracing techniques. Consistent with attribute weighting, eye-tracking studies looking at different choice biases have found that people tend to pay more attention to the attribute that the target item scores well on (Ashby, Dickert, & Glöckner, 2012; Noguchi & Stewart, 2014; Sütterlin et al., 2008; Balcombe, Fraser, Williams, & McSorley, 2017; Fisher, 2017). For example Balcombe et al. (2017) found that attributes with higher scores tended to be fixated on more, and these fixations partially predicted participants' preferences and their subjective ratings of attribute importance. While these eye-tracking results represent positive evidence for attribute weighting, it is unclear whether attention to an attribute is driving preferences or whether preferences are driving attention. It could be that some other mechanism is driving preference for a target item, which subsequently leads to increased attention to the attributes favouring that item, similar to the gaze cascade effect (Shimojo et al., 2003) or confirmation bias (Talluri, Urai, Tsetsos, Usher, & Donner, 2018; Nickerson, 1998).

One finding that attribute weighting cannot explain is Choplin and Hummel's (2005) finding that the attraction effect could occur even when options differed only along a single dimension. Specifically, rather than having participants make preferential choices, they had participants judge the similarity of options to a focal stimulus. These options differed along a single attribute, such as line length, where one option would have a higher score than the focal stimulus on the attribute and the other would have a lower score. They found that adding a decoy in that was similar to one of the choice options (the target) but further away from the focal stimulus (i.e. a dominated decoy) increased preference for the target option (i.e. an attraction effect was observed).

Though not reported as a test of attribute weighting, a similar result was observed by Fleming et al. (2010) for the default bias, wherein participants were more likely to judge that a ball fell inside of a line than outside when 'inside' was the default response and vice versa. While these tasks obviously represents a departure from the preferential choices we are interested in, it shows that the attraction effect and default bias generalise to conditions in which attribute weighting cannot apply. Findings that biases occur in conditions in which attribute weighting would not apply suggest that attribute weighting may not fully account for choice biases. Nevertheless, the ability to explain previously observed choice bias reversals along with positive eye-tracking results suggests attribute weighting may contribute to choice biases.

FIGURE 2.3: This figure is adapted from Choplin and Hummel (2005), showing an example of the stimuli used to show an attraction effect for choices that varied along a single dimension. The percentages represent the proportion of participants that chose the corresponding line as being most similar to the focal option.

2.3.5 Pre-decisional information distortion

Another means by which bias towards a target item could occur is due to decision makers evaluating the available information as favouring the target option and even coming up with additional reason to favour the target option (Russo, Meloy, & Medvec, 1998; Lord, Ross, & Lepper, 1979). This is often referred to pre-decisional information distortion, wherein the available information is distorted such that the target option seems more desirable than it may actually be, leading to preference for the target (Chaxel, Wiggins, & Xie, 2018; Russo, Medvec, & Meloy, 1996; Carlson, Meloy, & Russo, 2006)

This is most famously seen in the form of the confirmation bias, wherein people tend to evaluate evidence in such a way that they can maintain their current beliefs. For example, Lord et al. (1979) famously presented students with studies about the deterrent effect of the death penalty. They found that students who initially supported the death penalty interpreted these studies as favouring the death penalty and more strongly supported the death penalty as a result. Meanwhile, despite reading the same studies, students who initially opposed the death penalty more strongly opposed the death penalty after reading these studies. This finding that information can be interpreted differently to support pre-existing beliefs has been found to influence a wide array of judgements, ranging from simple perceptual judgements (Ounjai, Kobayashi, Takahashi, Matsuda, & Lauwereyns, 2018; Rajsic, Wilson, & Pratt, 2015) to complex criminal judgements (Nakhaeizadeh, Dror, & Morgan, 2014; Hergovich, Schott, & Burger, 2010; Lidén, Gräns, & Juslin, 2018). Given the apparent role that confirmation bias can have on our beliefs and judgements, it seems plausible that pre-decisional information distortion may also play a role in our preferential choices. This is most directly seen in the leader-driven primacy effect.

The leader-driven primacy effect refers to the finding that people tend to favour an option when early information favours that option over an alternative, even if subsequent information favours the alternative option (Russo et al., 1996; Russo et al., 1998; Carlson et al., 2006; Willemsen, Böckenholt, & Johnson, 2011). This was nicely exemplified by Carlson et al. (2006), who presented two options that varied on two attributes. Crucially, these attributes were presented serially, one after the other. Despite the same information being shown regardless of attribute ordering, Carlson et al. (2006) found that participants were biased towards choosing whichever option was favoured on the first-presented attribute. It was proposed that establishing an option as an early leader (through first showing an attribute that favoured that option over an alternative) led participants to process subsequent attributes more in favour of the early leader than they otherwise would. In support of this, Carlson et al. (2006) found that relatively neutral attributes (that did not clearly favour either option) were rated as more strongly favouring the early leader than the competitor option. This is consistent with the idea that information was being distorted to favour of the leader.

Typically, as seen in the examples above, pre-decisional information distortion is applied in contexts where people have an initial preference towards a certain option. With many choice biases, these explicit initial preferences are not necessarily present but may occur more subtly. For example, the inferior decoy in the attraction effect may establish the similar, superior target item as a leader relative to that decoy. The target option's superiority to the decoy option may be sufficient to cause information distortion in favour of it. A similar process may happen in the case of phantom decoy effects, where the superior phantom decoy encourages information distortion in favour of it and similar alternatives (i.e. the target item).

In the case of other biases such as the endowment effect, default bias, scarcity bias or norm effects, it may be that people engage in information distortion to justify why these manipulations have been applied to a particular option (Willemsen et al., 2011). For example, a descriptive norm may lead a decision maker to focus on advantages of the descriptive norm to justify that option's popularity, perhaps due to a meta-cognitive consistency goal of having all information consistently favour a single option (Simon, Krawczyk, & Holyoak, 2004; Festinger, 1962). One possible cause for this information distortion is a tendency to focus on evaluating certain information first, as proposed by Johnson et al. (2007) in explaining the endowment effect.

Johnson et al. (2007), in their query theory, propose that the endowment effect is caused by people's tendency to initially focus on reasons to keep their endowed item before coming up with reasons to not keep an endowed item. Due to memory effects such as the finding that recalling certain information (e.g. advantages of the endowed option) inhibits recall of other information (e.g. disadvantages of the endowed option; Dempster, 1995), they proposed that participants will come up with more reasons to keep their endowment. Thus, query theory proposes that people are biased towards considering more advantages of an endowed option, leading to the endowment effect. It is easy to see how this could be extended to other choice biases, by simply assuming that people first focus on the advantages of the scarce option, popular option, default option, exposed item etc.

To test query theory, Johnson et al. (2007) had participants evaluate endowed or non-endowed mugs by first listing aspects they were considering in making their decision and then assigning a monetary value to the mug. They found that participants who owned the mug listed more advantages of the mug than participants not endowed with the mug. Furthermore, they found that experimentally manipulating participants to list disadvantages of the endowed mug first could remove the endowment effect, although this did not go so far as resulting in a significant reversal of the endowment effect as query theory would predict. Thus, Johnson et al. (2007) at least partially support the role that pre-decisional information distortion may play in the endowment effect, though leave open the door for other mechanisms to also play a role.

In Section 2.2.4 we discussed the general lack of theory-testing studies for social norm effects and the tendency to explain these effects based on unique information offered by social norms (e.g. the social sanction account). In the context of lower-level, perceptual decision making though, there has been some more interest in the mechanisms underlying norm effects. These studies have offered support to a role of pre-decisional information distortion in norm effects. In a series of papers, Germar and colleagues (2014, 2016, 2019) looked at how social norms influenced decisions about visual stimuli by fitting computational models to data from seminal social influence paradigms. For example, Germar and Mojzisch (2019) looked at the mechanisms underlying social influence in the Sherif (1935) paradigm, wherein judgements about basic perceptual stimuli are influenced by knowledge of how other people responded.

The computational models they fit (specifically, diffusion models) had two separate parameters that could lead to a choice bias: shifts in the decision criteria such that decision makers were pre-disposed to conform to a norm and/or changing how perceptual information itself was processed, such that it favoured the norm (Smith, 2000; Ratcliff, Smith, Brown, & McKoon, 2016). While in this thesis we are not specifically interested in visual perception, it is possible that social norms affect visual perception in a similar way to how they affect preferential choice. What Germar and colleagues 2019, 2014, 2016 found when fitting these models is that the boundary shift parameter did not contribute to the model's ability to explain choice biases. Instead, the parameter representing biases in information processing changed based on manipulation of the social norm. Participants seemed to accumulate norm-congruent sensory information more than norm-incongruent information. Though these results are potentially dependent on the assumptions of the model being fit, were shown for perceptual decision making and are not exclusively explained by pre-decisional information distortion, they are consistent with the idea that people are biased toward interpreting and processing information such that it supports a social norm. Thus, it may be that the descriptive norm, along with various other choice manipulations, tend to encourage people to evaluate information in favour of the target option.

2.4 Conclusions

In this chapter, we have taken a broad look at many of the prominent mechanisms believed to underlie various choice biases. Though the rest of this thesis primarily focuses on just two of the biases discussed here (the endowment effect and the descriptive norm effect), having a broad understanding of the choice bias literature can guide our understanding of individual biases. Furthermore, we return to the theories discussed here throughout the thesis, discussing how theories from across the choice bias literature can potentially explain our empirical findings.

One trend that emerges in the literature is a general tendency to treat choice biases as highly distinct. This manifests not only in different choice biases tending to be studied under quite distinct contexts but also in an abundance of bias-specific theories. These bias-specific theories rely on the unique, salient aspects of the bias they attempt to explain. Adopting these bias-specific theories leads to a complex view of choice, which assumes fundamentally different mechanisms are engaged in different choice biases.

Through looking at the similarities between choice biases and potential overlap in their theoretical explanations, we can identify a number of generalisable mechanisms. These include loss and regret aversion, item attention, attribute weighting and pre-decisional information distortion. Such generalisable mechanisms have the potential to offer a more parsimonious account of a wide range of biases. For example, if endowments, defaults, scarce options, popular options etc. all tend to be established as reference points when making a decision, loss aversion could explain all of these biases. From a parsimony perspective, being able to explain all of these bases based on a general mechanism that acts across contexts should be favourable unless there is evidence to the contrary.

There is still a notable level of uncertainty as to how well these generalisable mechanisms explain different biases. For example, loss aversion remains a prominent explanation of many choice biases. However, we reviewed a finding by Brenner et al. (2007) that the endowment effect reversed (participants switched away from the endowment) when choosing between undesirable options. This finding is counter to the predictions of loss aversion, which assumes that losses relative to an endowed option will lead to endowment preference, even if options are undesirable. Instead, these results are consistent with item attention or attribute weighting theories, which propose that the value of the endowed option is amplified. In Chapter 3 and Chapter 4 we follow up on these findings to better understand the theoretical implications of such reversals.

We reviewed how, compared to other choice biases, our understanding of the mechanisms underlying social norm effects is distinctly underdeveloped. Not only have generalisable mechanisms not been applied to the descriptive and injunctive norm effects, but even the prominent, norm-specific theories have experienced minimal experimental testing. We discussed how this may be due to the fact that there are intuitive, objective reasons why people might follow a norm. However, given this lack of research, it is unclear whether people even are following social norms for these reasons. In fact, in Chapter 6 we show that people continue to follow descriptive norms even when there is no objective reason to conform to them. In Chapter 8 we find results that contradict the other prominent explanation of descriptive norm effects, self-categorisation theory. Having now reviewed various generalisable mechanisms across the choice bias literature, we discuss how these mechanisms could potentially explain our findings better than the prominent, norm-specific theories.

Chapter 3

Undesirable Choices: A Preface to Chapter 4

As noted in Chapter 2, loss aversion represents one of the most prominent and widely applied mechanisms for explaining a variety of choice biases. Some of the most compelling evidence against loss aversion, at least as an explanation for the endowment effect, comes from Brenner et al.'s (2007) finding that the endowment effect reverses when people are choosing between two undesirable options. Brenner et al. (2007) found that participants preferred to switch away from an undesirable endowed option in favour of an equally-undesirable alternative. This finding was taken as evidence that the endowment effect is not driven by loss aversion because loss aversion considers the absolute value of options to be irrelevant during the endowment effect. Instead only the value of options relative to a reference point (e.g. the endowed option) are relevant for loss aversion, and these relative values are the same regardless of whether the options are both desirable or undesirable. Though not explicitly testing loss aversion, similar reversals have been found for serial order (Houston et al., 1989) and mere exposure (Armel et al., 2008) effects.

While Brenner et al. (2007) discuss the implications of their findings for loss aversion, they also may be relevant for a range of different theories. In the introduction, we discussed a number of general mechanisms that can potentially explain a wide range of choice biases. Some of these mechanisms also predict that choice biases should remain consistent, regardless of whether the choice is between desirable or undesirable options. For example, pre-decisional information distortion proposes that we tend to focus on and think about the positive aspects of the target item and the negative aspects of the alternative item (Johnson et al., 2007; Willemsen et al., 2011; Russo et al., 1996). Like loss aversion, pre-decisional information distortion is not thought to depend on the absolute value of the options and so would not predict choice biases to reverse when choosing between undesirable options. This prediction is also made by regret aversion and all of the bias-specific mechanisms that were discussed in Chapter 2, as reviewed in Appendix A on page 143.

In contrast, some theories do predict a reversal of biases for choices between undesirable options. Attribute weighting theories such as salience theory and the associative accumulation model, for example, propose that we weight attributes based on their activation (Bhatia, 2013; Bordalo et al., 2012a, 2012b). Activation of an attribute is said to be determined by the prevalence of that attribute amongst items, weighted by the salience of each item. So, for example, an endowed item might be particularly salient, leading to the attributes it scores highly on receiving greater activation and thus, being weighted more in the decision. Crucially, this means that if the endowed item scores highly on a negative attribute, this attribute will tend to get weighted more strongly than attributes that the alternative options score highly on. When the target item scores highly on a negative attribute, increased weighting of this attribute is predicted to decrease preference for the target item.

Like attribute weighting, item attention theories also predict choice biases to reverse for choices between undesirable options. Increased attention towards a target item is thought to result in evidence accumulating in favour of the target item if it is desirable and against the target item if it is undesirable (Krajbich, Armel, & Rangel, 2010; Shimojo et al., 2003; Armel et al., 2008). So, for example, if an undesirable endowed item tends to attract greater attention than an undesirable competitor, this will accumulate more evidence against the endowed option such that the decision maker wants to switch away from it (an endowment reversal).

Given these theoretical implications, we were particularly interested in looking further into Brenner et al.'s (2007) endowment reversal. The initial impetus for following this line of research was to see if similar reversals occurred for a range of other choice biases. If other choice biases reverse when choosing between undesirable options, this would suggest that these biases might be driven by mechanisms such as item attention or attribute weighting, which predict such reversals. Meanwhile, finding that some biases do not reverse would indicate that they rely on meaningfully different mechanisms to those that do reverse, such as loss aversion, regret aversion or pre-decisional information distortion.

To test whether reversals for undesirable choices occurred across a range of choice biases, we ran a pilot study that looked at a number of prominent biases for both desirable and undesirable choice scenarios.

3.1 Method

3.1.1 Participants

85 participants (Mean age = 30, 31% female) were recruited through Microworkers.com and paid $5.00 for participating. Participants remotely completed the experiment in a web browser after providing informed consent. The experiment took around 40 minutes to complete.

3.1.2 Procedure

After providing some demographic information, participants were provided with instructions for the task. They then answered the following questions about these instructions to ensure they understood the task. If they answered any questions incorrectly, they were returned to the instructions page and to reread the instructions before attempting the understanding check again.

- Your objective in this experiment is to:
  1. choose the option that is the most attractive to you based on the information displayed. *(correct)*
  2. choose the option that is the least attractive to you based on the information displayed.
  3. choose any option by randomly making a choice.
- Is the following statement true or false? "On some, but not all, trials there will be a clearly correct choice."
  1. True; if one option is rated better than all others by both estimates then this option is clearly preferable and should be chosen. *(correct)*
  2. False; all options will be approximately equally attractive and there is never a clearly correct option.
- Is the following statement true or false? "On some attributes, higher values are better. On other attributes, lower values are better."
  1. True; for example, higher values on quality are good whilst higher values on price are bad. *(correct)*
  2. False; the attributes have been chosen such that lower values are always better.
- Is the following statement true or false? "Though only one attribute is presented, the options may differ on a range of other attributes that are not shown."
  1. True; because of this, you should not regard the displayed values too strongly.
  2. False; the options should be assumed to be equivalent on any attributes not presented. *(correct)*
- Is the following statement true or false? "Some options will appear equally attractive - in such cases I should just go with my personal preference."
  1. True; I should simply choose the option that appears most attractive to me. *(correct)*
  2. False; This is when I ask a friend or run a search on Google to find what the best option actually is.
- Is the following statement true or false? "The attribute estimates on one trial are related to the attribute estimates on previous trials."

1. True; it is therefore important to remember the previously presented attribute values when making your choices.
2. False; each trial is self-contained so you should forget about any previous trials when making your choice on the current trial. *(correct)*

After answering all of these questions correctly, participants began the main experimental phase, which involved making a series of choices between options (see Figure 3.1 on the next page). Each choice involved two key options (in the case of decoy effects, an additional decoy option was presented). The options differed along estimates from two reviewers of a single attribute. For the sake of simplicity, we refer to each reviewers estimates as distinct attributes.

The reviewer attributes were randomised within certain ranges with the restriction that the average of the reviewer's estimates be equal. This allowed us to ensure that both choice options were approximately equally desirable. However, for 20% of trials, the attribute estimates were changed so that both reviewers favoured the same option. These catch trials offered an opportunity to ensure that participants were engaging with the task. Seven participants from the undesirable condition and ten participants from the desirable condition were excluded from the analysis because they incorrectly chose the inferior option (the option rated lower by both reviewers) on 25% or more of these catch trials.

For approximately half (44) of the participants, these choices were between desirable options. Specifically, these participants were presented with a series of choices, with each choice involving options taken from one of the following choice categories:

- Printers that varied in estimated printing speed
- Laptops that varied in estimated battery life
- Bikes that varied in estimated weight
- Washing machines that varied in estimated water consumption
- Cameras that varied in estimated picture quality

For the remaining participants (41), these choices were between undesirable options. Specifically, the choice categories were:

- Travel locations that varied in the estimated severity of a rain storm
- Delayed flights that varied in the estimated length of delay
- Medications that varied in estimated risks of side effects
- Factories to produce a product in that varied in their estimated manufacturing cost.
- Undesirable jobs that varied in estimated commute length

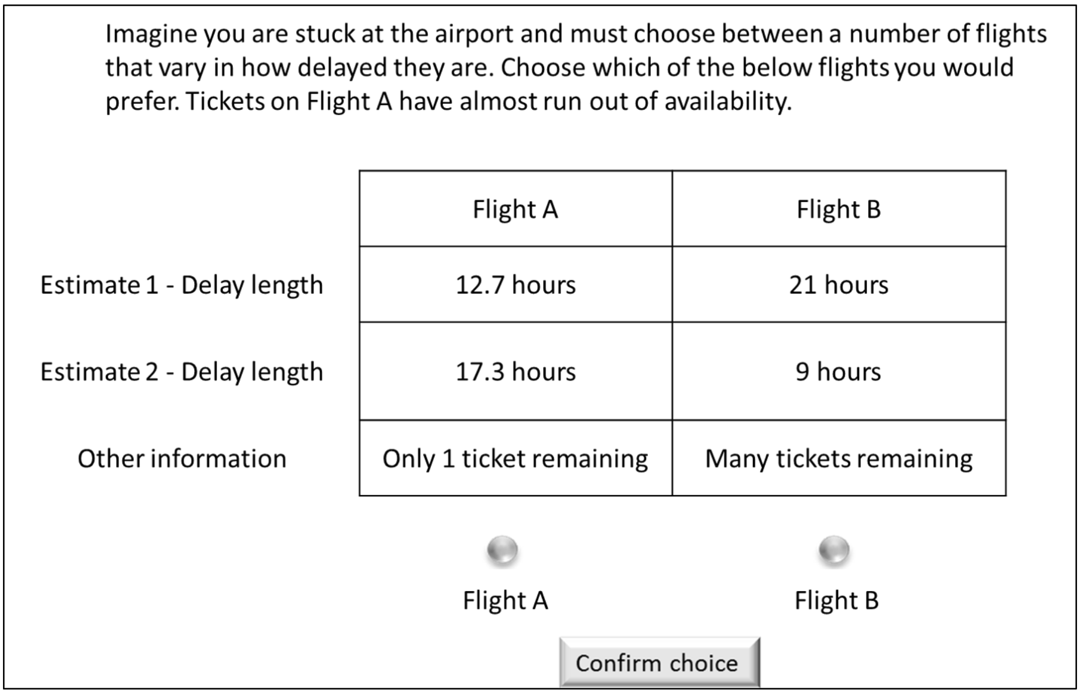

FIGURE 3.1: Example of a trial involving choice between two delayed flights. This particular example involves a scarcity bias manipulation where Flight A (the target option) is said to have limited availability.

Each choice involved a manipulation intended to induce one of 10 different choice biases as outlined below. Below we outline how each choice manipulation would appear for the particular example in which participants were choosing between two delayed flights, where Flight A was the target option and Flight B was the competitor option.

Default bias Flight A was selected by default. Text additionally stated "Flight A is the default option but you may switch this for Flight B."

Endowment effect Text stated that "You own a ticket on Flight A. This is your flight. You have the option to switch your flight for Flight B or keep Flight A." Below the attribute information for Flight A was the text "Owned" while the text "Not owned" was below Flight B.

Mere exposure effect Flight A was presented on the screen with the corresponding attribute information. Participants had to click a button to reveal information about both flights and make their choice.

Leader-driven primacy effect Participants were told that "Information about each estimate will be presented separately. Read the current estimate information carefully because it won't be presented again." The attribute that Flight A scored better on was presented first. After clicking a button, this attribute information was hidden and information about the other attribute (which favoured Flight B) was presented. After another button click, all attribute information was hidden and participants chose between the two options.

Scarcity bias Flight A was said to have "Only 1 ticket remaining". Flight B was said to have "Many tickets remaining".

Descriptive norm effect Participants were informed that "Flight A has been chosen by most people completing this experiment". Below the attribute information for Flight A was the text "More popular". Below Flight B was the text "Less popular".

Injunctive norm effect Participants were told that we recommended they choose Flight A. Below Flight A was the text "Recommended MOST for you". Below Flight B was the text "Recommended LEAST for you".

Attraction effect A third decoy alternative was added that was similar to one of the options but had a score 10% worse score on one attribute

Unavailable attraction effect Same as the attraction effect except that the decoy was unavailable (the option to choose it was greyed out and text stated the option was unavailable).

Phantom decoy effect Same as the unavailable attraction effect except that the unavailable decoy option was 10% better than a target option on one attribute.

3.1.3 Design

A mixed-design was used. Whether the choice stimuli were desirable or undesirable varied between-subjects while the choice bias presented, which option was the target and various other features of the choice stimuli were varied within-subjects.

The experiment consisted of 10 blocks of trials. In each block a different manipulation was made to try to introduce a different choice bias (e.g. one block involved an option being set as the default on each choice trial, another block involved attraction decoys being presented for each choice trial etc.). For each block, 25 trials were presented, consisting of five trials for each of the five different choice stimuli (e.g. participants in the undesirable condition made five choices between various risky medicines, five choice between delayed flights etc.). For each type of choice stimuli in each block, one trial was designated as a catch trial, where the value of one option was superior to the other option on all attributes such that it should be chosen. 10 participants in the desirable condition and 7 participants in the undesirable condition who failed to choose the superior option on at least 75% of trials were excluded from analysis (although the results remain qualitatively the same when they are included). The order of blocks and the order of choice categories were randomised with the restriction that participants would see one of each choice category before any choice category could be repeated.

Within each block, various aspects of the target were controlled for such as which attribute it was preferred on, the position it was presented in the table (left or right) and whether the reviewers had more differing ratings (extreme target) or more similar ratings (consistent target) for the target as compared to the competitor. This meant that a null effect of the manipulation should result in a target choice share of approximately 50%, even if participants were biased towards a particular attribute, the left/right option or more/less extreme ratings. Thus, the percentage of total trials in which the target was chosen over the competitor provides an estimate of the size of each choice effect, with significant deviations from 50% indicating significant effects of the manipulation.

3.2 Results and discussion

For each choice bias, we collapsed across all trials and participants to look at the relative proportion of trials in which a target option was chosen over a competitor option. We failed to observe reversals for any of our choice biases (see Figure 3.2 and Table 3.1 on the next page). Instead, every choice bias that showed a significant effect was in the same direction (favouring the target option) regardless of desirability.

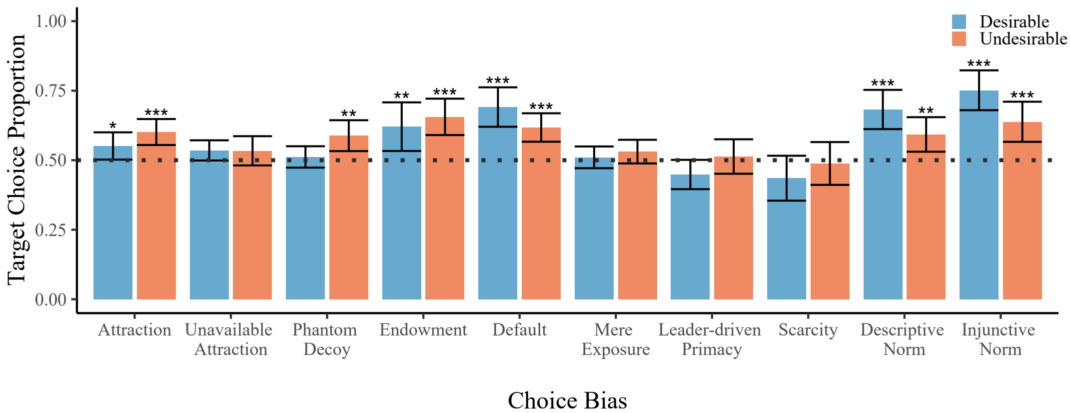

FIGURE 3.2: Bar chart showing the relative proportion of trials in which the target option was chosen over a competitor option for a range of choice biases from the pilot study described in this chapter. Blue bars represent choices between desirable options while red bars represent choices between undesirable options. The error bars show 95% confidence intervals and asterisks show significant p values ($*\,*\,*$: $p < .001$, $**$: $p < .01$, $*$: $p < .05$). It can be seen that target choice proportion never drops significantly below 0.5, meaning that none of the choice biases showed evidence of reversing due to desirability.

Crucially, we found this even to be the case for the endowment effect, with participants preferring to keep the endowed item even when it was undesirable. This raised the question of why our results differed so markedly from Brenner et al.'s (2007). Thus, instead of further studying how a broad range of choice biases were affected by undesirable options, Chapter 4 takes a deeper look at endowment reversals, offering insights into when and why they occur.

One notable limitation of this pilot study is that the categories of choice stimuli in the desirable condition differed from those in the undesirable condition. Thus,

TABLE 3.1: Statistics for each condition in the pilot study

Effect	**Desirability**	**Target choice proportion**	**95% CI**	***p* value**
Attraction	Desirable	55.1%	[50.2%, 60.0%]	.041
	Undesirable	60.1%	[55.5%, 64.8%]	.0001
Unavailable Attraction	Desirable	53.5%	[49.9%, 57.2%]	.057
	Undesirable	53.4%	[48.1%, 58.6%]	.199
Phantom Decoy	Desirable	51.2%	[47.3%, 55.0%]	.538
	Undesirable	58.8%	[53.3%, 64.4%]	.003
Endowment	Desirable	62.1%	[53.3%, 70.8%]	.008
	Undesirable	65.6%	[59.0%, 72.1%]	<.0001
Default	Desirable	69.1%	[62.0%, 76.2%]	<.0001
	Undesirable	61.8%	[56.6%, 66.9%]	<.0001
Mere Exposure	Desirable	51.0%	[47.1%, 54.9%]	.596
	Undesirable	53.1%	[48.8%, 57.3%]	.149
Leader-driven Primacy	Desirable	44.9%	[39.6%, 50.1%]	.055
	Undesirable	51.3%	[45.1%, 57.5%]	.667
Scarcity	Desirable	43.5%	[35.5%, 51.6%]	.113
	Undesirable	48.8%	[41.1%, 56.5%]	.758
Descriptive Norm	Desirable	68.2%	[61.2%, 75.3%]	<.0001
	Undesirable	59.3%	[53.1%, 65.5%]	.005
Injunctive Norm	Desirable	75.1%	[68.0%, 82.3%]	<.0001
	Undesirable	63.8%	[56.6%, 71.0%]	.0004

had we observed differences between the desirable and undesirable conditions, it would be unclear if this was due to valence (desirable or undesirable) or due to the choice categories. While we have no a priori reason to believe the particular choice categories used in the desirable and undesirable conditions would have a systematic impact on choice biases, the experiments in the next chapter ensure that the choice categories are equivalent in the desirable and undesirable conditions. This allows us to directly compare preference for the endowed option in the desirable and undesirable conditions without being confounded by changes in the choice category.

Chapter 4

Reversing the Endowment Effect

The contents of this chapter are adapted from our paper published in *Judgement and Decision Making*: Pryor, C. G., Perfors, A., & Howe, P. D. L. (2018). Reversing the endowment effect. *Judgement and Decision Making*, *13*(3), 275–286

4.1 Abstract

When given a desirable item, people have a tendency to value this owned item more than an equally-desirable, unowned item. Conversely, when the endowed item is undesirable, in some circumstances people have a tendency to swap it for an equally undesirable item, a phenomenon known as the reversed endowment effect. The fact that the endowment effect can reverse for undesirable items has been taken as evidence against loss aversion being the underlying cause of the endowment effect. This study represents the first time that the reversed endowment effect has been observed for choices with real consequences. However, we find that the reversed endowment effect occurs only when participants' ability to compare the available choice options is limited. We further show that these endowment reversals can also be induced for choices between desirable options and removed for choices between undesirable options by manipulating the expectations participants have when making a choice. Finally, we show that our data, including endowment reversals, can in principle be explained by loss aversion.

4.2 Introduction

The endowment effect, the finding that mere ownership of an item tends to increase the value assigned to it, has maintained consistent research interest for decades (Thaler, 1980). Valuation paradigms have found that people often require more money to give up ownership of an object than they would be willing to pay to acquire that same object (Kahneman et al., 1990). Furthermore, preferential choice paradigms have found that preference for an item tends to be greater when it is owned than when an alternative item is endowed (Knetsch, 1989). Interestingly, Brenner et al. (2007) found that this preference for keeping endowed items reversed when the items were undesirable.

Brenner et al. (2007) presented participants with hypothetical choices between two undesirable options. For example, they told participants to imagine they had received a speeding ticket and had been endowed with paying a monetary fine for that ticket. Participants were offered the option to exchange their endowed fine for an alternative punishment: attending traffic school. The standard endowment effect would suggest that people should prefer to keep their endowed fine more so than if they had instead been endowed with attending traffic school. Counter to this, Brenner et al. (2007) found that participants were more likely to switch away from their undesirable endowed option than keep it.

Brenner et al.'s (2007) results are particularly interesting from a theoretical perspective because they argue against the most prominent explanation of the endowment effect, prospect theory. Prospect theory, and related theories, propose that we tend to evaluate options relative to our current reference state (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992; Kőszegi & Rabin, 2006). When endowed with an option, it is proposed that an individual's new, endowed state is set as their reference point so that they evaluate all other options relative to the endowed option. An equivalently-valuable, non-endowed alternative will have some advantages (gains) and some disadvantages (losses) relative to the endowed option. Importantly, prospect theory assumes that people are loss averse such that the relative losses of the non-endowed alternative loom larger than that alternative's relative gains. Thus, the individual is assumed to be more influenced by the potential losses of switching to the non-endowed option, than the potential gains, leading to preference for the endowed option so as to avoid these losses.

As Brenner et al. (2007) point out, prospect theory only focuses on the gains and losses of an alternative relative to the endowed option. Thus, according to prospect theory, only the relative value of the endowed and alternative option should be relevant to which option is preferred, not their absolute value. This means that whether the options are both desirable or both undesirable should be irrelevant. As a result, under these assumptions, prospect theory cannot explain Brenner et al.'s (2007) endowment effect reversals for undesirable options.

An explanation of the reversed endowment effect, proposed by Brenner et al.

(2007), revolves around the idea that *all* attitudes towards the endowed option are amplified. When both options are desirable, the desirability of the endowed option is amplified, leading to the standard endowment effect. When both options are undesirable, the undesirability of the endowed option is amplified, leading to an endowment reversal. Brenner et al. (2007) suggested that this amplification occurs because losing the endowed option itself is weighted more than gaining an alternative item. Thus, losing a desirable option that you own is particularly painful (explaining the endowment effect) whilst losing an undesirable option that you own is particularly positive (explaining the endowment reversal). Attentional theories such as salience theory and the associative accumulation model have a similar interpretation, suggesting that we amplify the valence of the endowed option it is more salient, making the desirable or undesirable features of the endowed option stand out more (Bordalo et al., 2012a, 2013; Bhatia, 2013).

Whilst Brenner et al.'s (2007) observed endowment reversals have been taken as support for this amplification effect, recent research suggests the story is not so simple. Dertwinkel-Kalt and Köhler (2016) looked at whether the reversed endowment effect, which was previously observed for hypothetical choices, generalised to incentivised choices. Participants chose between tedious, undesirable tasks that they subsequently either had to complete (incentivised condition) or did not have to complete (hypothetical condition). Under the hypothetical condition, they found data consistent with Brenner et al.'s endowment reversal; participants preferred to switch away from their endowed task. However, when the choices were incentivised they found that participants returned to a standard endowment effect, preferring the endowed task. While amplification effects can explain an endowment reversal in the hypothetical condition, they cannot explain the preference for undesirable endowments in the incentivised condition.

Dertwinkel-Kalt and Köhler (2016) explained their observed difference between incentivised and hypothetical choices with a dual-process model. They suggested that, in line with salience theory, the endowed option is processed first. For hypothetical choices, people are not motivated to engage in any further processing. This leads to favouring desirable endowments or avoiding undesirable endowments in the manner predicted by salience theory. However, when participants are sufficiently incentivised, they will engage in the more effortful process of comparing the options. These direct comparisons engage loss aversion, which then shifts preference towards the endowed option regardless of desirability.

Dertwinkel-Kalt and Köhler's (2016) explanation of their results suggests that endowment reversals should tend to occur when one's ability to process and compare the choice stimuli is limited. The first aim of this paper is to explicitly test this prediction by limiting comparisons between the choice options during incentivised choices. We do this in two separate ways. First, by hiding information about the non-endowed alternative to explicitly replicate the lack of processing of the non-endowed options hypothesised by Dertwinkel-Kalt and Köhler. Second, by enforcing response deadlines in order to limit the time available to compare the choice options. If limited comparisons between the options are important to the occurrence of endowment reversals, then the endowment effect should reverse between desirable and undesirable choice settings under either of these limited-comparison conditions while no such reversal should occur when comparisons are freely allowed. To preempt our results, we indeed found that endowment reversals occurred only when comparisons were limited.

The second aim of this paper is to test an alternative explanation of the reversed endowment effect. If the reversed endowment effect is dependent on comparisons with the non-endowed alternative being limited, an interesting implication is that these reversals could be due to people evaluating the endowed option relative to their expectations rather than due to amplification of the endowed item. For example, people have a tendency to not update their beliefs and expectations entirely, when presented with new information (Anderson, 1983; Nissani & Hoefler-Nissani, 1992). Thus, people may not perfectly update their reference point when endowed with an item and instead, continue to be influenced by the relatively neutral reference state they had prior to being presented with the choice scenario. When comparisons with the non-endowed alternative are limited, desirable endowments would exceed these more neutral expectations and therefore be preferred (the endowment effect) while undesirable endowments would be inferior to these more neutral expectations and therefore tend to be avoided (the reversed endowment effect).

This explanation of the reversed endowment effect based on imperfect updating of the reference point makes an interesting prediction. Previously, reversals of the endowment effect were attributed to the choice options being undesirable. However, the theory we present here suggests that endowment reversals are not based on the absolute desirability of the options but instead on their value relative to one's previous reference point. This leads to the prediction that endowment reversals will occur both for desirable and undesirable choices if someone had a previous reference state that is superior to the endowed option. Conversely, it is predicted that the endowed option will be preferred, regardless of whether it is desirable or undesirable, if someone had a previous reference state that was inferior to the subsequently endowed option.

In Experiment 4 we tested this prediction using a practice gamble to induce a superior or inferior reference point prior to the choice of interest. We found that regardless of whether the endowed option was desirable or undesirable, the endowment effect would reverse if the reference point was superior to the endowed option but would not reverse if it were inferior to the endowed option. We also found that a model based on prospect theory that incorporates these assumptions could explain our data. This shows that endowment reversals cannot, in principle, be taken as evidence against prospect theory, as has previously been claimed. Of course, this does not mean that prospect theory is necessarily the best explanation of our data and to emphasise this point, we concluded by considering possible alternative explanations for our data.

4.3 Experiment 1

This experiment directly tested the hypothesis that whether the endowment effect reverses or not is influenced by the extent to which the two options are compared (Dertwinkel-Kalt & Köhler, 2016). It did this by sometimes showing the non-endowed alternative (thereby allowing it to be compared to the endowed option) but other times hiding it so as to prevent comparisons. This design thus forced participants to process the stimuli in the manner that the dual-process model of Dertwinkel-Kalt and Köhler (2016) assumed occurs for hypothetical vs incentivised choices. Specifically, we hypothesised that participants would tend to favour the endowed option over the non-endowed option regardless of desirability when making hypothetical choices. On the other hand, we hypothesised that when the choice was incentivised and the choice stimuli were undesirable, participants would tend to favour the non-endowed option.

We used risky gambles, traditionally referred to as "prospects", as the choice stimuli. This was done for two reasons. First, it allowed us to present incentivised choices by informing participants that the outcome of the gambles would influence their payment. Second, with gambles it is easy to create directly equivalent desirable and undesirable choices. Previous studies have used undesirable scenarios that do not have a clear equivalent desirable scenario. For example, the undesirable tasks used by Dertwinkel-Kalt and Köhler (2016) consisted of sorting confetti or filling a grid with '1's and '0's. For neither of these tasks is there an obvious equivalent desirable task to compare to. Gambles allow desirable options (a probability of winning a certain amount) to be precisely matched to equivalent undesirable options (the same probability of losing the same amount). Finally, it is worth noting that we presented choices between two gambles rather than one gamble and a non-risky (certain) alternative simply to reduce the biases people have been found to show towards increasingly certain outcomes (Kahneman & Tversky, 1979), which might otherwise have overpowered any effect of the endowment manipulation.

Given that the choices we presented were incentivised, the dual-process model proposed by Dertwinkel-Kalt and Köhler (2016) predicted that the endowed option would be preferred for both desirable and undesirable gambles when comparisons were allowed. When comparisons were prevented by hiding information about the non-endowed gamble, it was predicted that a regular endowment effect would be observed for desirable gambles, but a reversed endowment effect would be observed for undesirable gambles.

4.3.1 Method

200 participants (mean age = 32 years old, 51% female) were recruited online through Microworkers.com and completed the experiment in a web browser. All participants in all experiments presented her years olde were from English-speaking countries and provided informed consent before participating. The task was advertised as paying US$0.20 plus a potential additional bonus. However, upon completion, all participants were paid $0.80, corresponding to the best possible outcome that could be expected based on bonuses offered during the experiment. The experiment took around 2 minutes to complete.

Upon beginning the experiment, participants were informed that they were currently earning $0.50 for participation in the experiment but had to choose a gamble which could affect their payout. The gambles could win them up to $0.30 or lose them up to $0.30 (reducing their payout to as low as $0.20). Participants were presented with a screen that was completely empty except for two boxes labelled "Gamble A" and "Gamble B" respectively. They were asked to choose one of the boxes before any information about the gambles was presented. Whichever box they chose was designated as the endowed gamble.

After choosing a box, participants were then shown a gamble and endowed with this gamble. To be clear, their choice of box was, in reality, irrelevant because whichever box they chose was then assigned the values associated with the endowed gamble allocated to them. The endowed gamble was chosen from the following two gambles with equivalent expected value: 40% chance of $0.30 or 80% chance of $0.15, such that half of participants were endowed with the former gamble and half were endowed with the latter. In the DESIRABLE condition, participants were told they had a chance of winning the stated amount. In the UNDESIRABLE condition, they were told they had a chance of losing the stated amount. Participants then clicked a button to continue.

If the participant was in an ALLOWED comparison condition, then they were now shown the non-endowed gamble alongside the endowed gamble. If the endowed gamble was a 40% chance of $0.30 then the alternative gamble was an 80% chance of $0.15, and vice versa. Alternatively, if the participant was instead in the HIDDEN condition then the alternative gamble was not shown and instead was represented by a question mark, making them unable to compare the endowed gamble to the non-endowed alternative. The dependent variable of interest was whether participants then chose to keep the endowed gamble or swap to the non-endowed gamble.

These conditions amounted to a 2x2x2 between-subjects design crossing **desirability** (DESIRABLE or UNDESIRABLE) x **comparisons** (ALLOWED or HIDDEN) x **endowed gamble** (40% of $0.30 or 80% of $0.15). The latter variable was included to control for possible biases in the sample such as risk seeking or risk aversion, and we therefore collapsed across it.[1] Through controlling for such biases, we could expect

[1] A chi-square test looking at endowment preference based on which option was endowed found no significant difference in overall endowment preference between when the 40% chance of $0.30

that if participants were unaffected by the endowing of an item then **endowment preference** (the percentage of participants choosing the endowed item) should be approximately 50%. Endowment share greater than 50% would reflect preference for the endowed gamble over the non-endowed alternative.

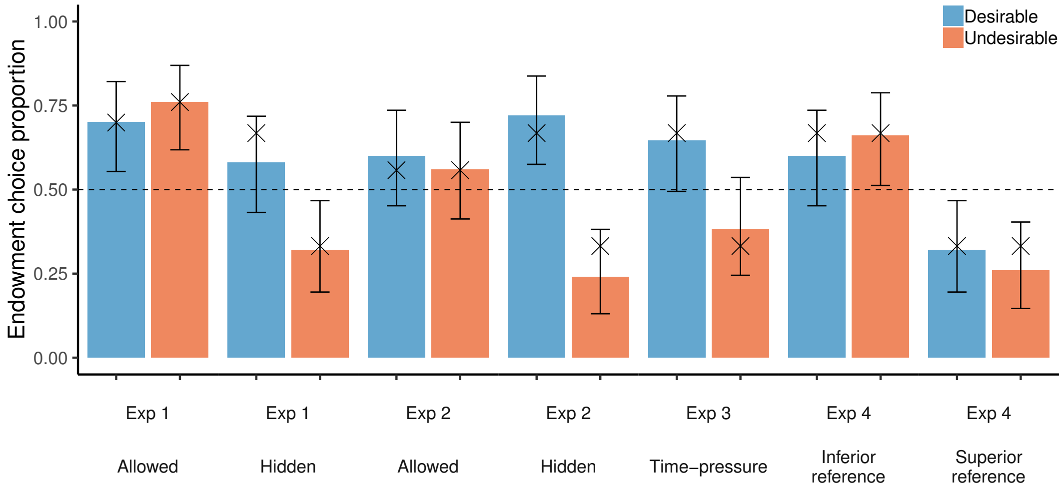

FIGURE 4.1: Bar chart of observed endowment preference share (error bars represent 95% CIs) across the different conditions of the experiments reported in this chapter. The crosses (Xs) represent the predicted **endowment preference** when the model of loss aversion described later in this paper was simultaneously fit to our different experimental conditions.

4.3.2 Results

The average **endowment preference** and 95%CI for each condition is shown in Figure 4.1. How were people's choices affected by the desirability of the options and whether comparisons were possible? We addressed this question with a 2x2 logistic regression with **desirability** (UNDESIRABLE = 0, DESIRABLE = 1) and **comparisons** (HIDDEN = 0, ALLOWED = 1) as predictors and **endowment preference** (non-endowed alternative chosen = 0, endowment chosen = 1) as the outcome variable. There were significant main effects of both **desirability** ($OR = 2.93$, $p = .013$, $95\%CI[1.31, 6.77]$) and **comparisons** ($OR = 6.73$, $p < .001$, $95\%CI[2.96, 16.78]$). Importantly, there was also a significant interaction ($OR = 0.25$, $p = .025$, $95\%CI[0.07, 0.83]$) with the endowed gamble being significantly more likely to be chosen in the DESIRABLE condition compared to the UNDESIRABLE condition when the non-endowed alternative was HIDDEN ($\chi^2(1, 100) = 6.83$, $p = .009$, $OR = 2.93$, $95\%CI[1.30, 6.65]$). On the other hand, in the ALLOWED condition, no significant difference was found between the DESIRABLE and UNDESIRABLE conditions ($\chi^2(1, 100) = 0.46$, $p = .50$, $OR = 0.74$, $95\%CI[0.30, 1.79]$). Consistent with the proposition of Dertwinkel-Kalt

gamble or the 80% chance of $0.15 gamble was endowed ($\chi^2(1, 200) = 0.33, p = .565, OR = 1.18, 95\%CI[0.67, 2.07]$).

and Köhler (2016), the desirability of the gamble affected choices only when the comparisons were hindered. Experiment 2 looks at whether these results generalise to non-risky choice stimuli.

4.4 Experiment 2

Experiment 2 tests whether the results of Experiment 1 can be replicated for non-risky choice stimuli. There is evidence that people respond differently to desirable risks than they do to undesirable risks (Kahneman & Tversky, 1984), raising the possibility that our observed effect of desirability on hypothetical choices in Experiment 1 was due to the use of risky prospects. Nevertheless, we hypothesised that participants would continue to show preference for the non-endowed option in undesirable hypothetical choice scenarios even when the choice options were riskless.

4.4.1 Method

200 English-speaking participants (mean age = 35 years old, 53% female) were recruited online through Amazon's Mechanical Turk. The design of Experiment 2 was exactly the same as Experiment 1 except for the changes outlined below.

Participants were informed that they would have to complete a button-pressing task during this experiment. Participants in the DESIRABLE condition were initially informed that they would "be paid \$0.40 for completing the task on the next page, which takes about 1.5 minutes and requires clicking on very small buttons". However, they were then asked to choose one of two boxes which might change the task involved or the monetary payout. For half of the participants in the DESIRABLE condition, their chosen box endowed them with a "20 second task instead, which requires clicking on large buttons". The other half of participants in the DESIRABLE condition were endowed with an additional \$0.10 for completing the experiment. We collapsed across these endowed gamble conditions when analysing the data to control for possible biases towards a shorter task or higher payout affecting **endowment preference**.[2]

Participants in the ALLOWED comparison condition were shown the other possible option (e.g. participants endowed with an additional \$0.10 were shown that the other box offered the 20 second task). As in Experiment 1, all participants were then informed that they had the option to switch to the other box if they wished and asked to choose which they would prefer.

The procedure for participants in the UNDESIRABLE condition was exactly the same and had the same potential end-points as those in the DESIRABLE condition, but the available options were phrased as being undesirable. Specifically, participants were initially informed that they would be "paid \$0.50 for completing the task on the next page, which takes about 20 seconds and involves clicking on large

[2]Whether the shorter task or higher payout was endowed did not significantly affect **endowment preference** ($\chi^2(1, 200) = 0.32$, $p = .57$, $OR = 1.17$, $95\%CI[0.67, 2.05]$).

buttons". The two boxes either endowed participants with completing "a tricky 1.5 minute task instead, which requires clicking on very small buttons" or a reduction in their payout of $0.10.

Regardless of their allocated condition or decision, upon deciding which option to keep, participants were presented with a debriefing statement and informed that the choice was actually only hypothetical and that they did not have to complete any additional task. They were paid $0.50 for participation, which took around 1 to 2 minutes.

4.4.2 Results

Figure 4.1 on page 47 shows the average endowment preference and 95% CI for each experimental condition in Experiment 2.

A 2x2 logistic regression found a significant main effect of **comparisons** ($OR = 4.03$, $p = .001, 95\%CI[1.75, 9.76]$) and of **desirability** ($OR = 8.14$, $p < .001, 95\%CI[3.42, 20.69]$) on **endowment preference**. Consistent with the finding from Experiment 1, a significant interaction between **desirability** and **comparisons** was observed ($OR = 0.14$, $p = .002$, $95\%CI[0.04, 0.47]$). Specifically, the endowed option was chosen significantly more in the DESIRABLE than the UNDESIRABLE condition for the HIDDEN comparison condition, ($\chi^2(1, 100) = 23.08$, $p < .001$, $OR = 8.14$, $95\%CI[3.32, 19.94]$), but not for the ALLOWED comparison condition ($\chi^2(1, 100) = 0.16$, $p = .685$, $OR = 1.18$, $95\%CI[0.53, 2.61]$). Thus, our results from Experiment 1 replicated with non-risky choice options.

4.5 Experiment 3

Experiments 1 and 2 limited comparisons by hiding the non-endowed alternative. This manipulation represents the most extreme extent to which participants could fail to compare options when making a choice. Experiment 3 looks at whether the findings from the previous experiments generalise to a setting in which both options are presented, and comparisons are instead limited simply by enforcing a response deadline. It was hypothesised that participants would favour the endowed option regardless of desirability when they were able to take their time making a choice. On the other hand, when decision time was limited through a response deadline, we hypothesised that participants would favour the non-endowed option during undesirable choice scenarios.

4.5.1 Method

100 English-speaking participants (mean age = 36 years old, 50% female) were recruited online through Amazon's Mechanical Turk. Experiment 3 used the same design as the DESIRABLE and UNDESIRABLE ALLOWED comparison conditions of

Experiment 1 except participants were now instructed at the start of the experiment that a countdown timer might appear during the experiment.[3] To encourage rapid responding, they were instructed that failure to make a choice before the timer reached zero might reduce their potential bonus payout. When participants were offered the option to swap their endowed gamble with the non-endowed alternative, a timer always appeared on the screen counting down from ten seconds.

Five participants (three from the UNDESIRABLE condition and two from the DESIRABLE condition) were excluded from the analysis due to failing to respond within the ten second deadline. These participants were still paid the full bonus for participating. The conclusions drawn from this experiment remained the same when these participants were included in the following analyses.

4.5.2 Results

Figure 4.1 on page 47 shows the mean **endowment preference** share and 95%CI for the DESIRABLE and UNDESIRABLE conditions.

To test the prediction that there should be a greater difference between desirable and undesirable choices in terms of endowment preference when decisions are made under time pressure, we compared responses in Experiment 3 (the TIME-PRESSURE condition) to the ALLOWED comparison condition of Experiment 1. A 2x2 logistic regression predicted **endowment preference** based on **desirability** and **comparisons** (TIME-PRESSURE = 0, ALLOWED = 1). A significant main effect of **comparisons** was found ($OR = 5.1$, $p < .001$, 95%$CI[2.17, 12.63]$), though not of **desirability** ($OR = 0.74$, $p = .500$, 95%$CI[0.30, 1.79]$). Of key interest, there was a significant interaction between **desirability** and **comparisons** ($OR = 0.25$, $p = .026$, 95%$CI[0.07, 0.84]$) with the endowed gamble being significantly more likely to be chosen in the DESIRABLE condition compared to the UNDESIRABLE condition when the choice was made under TIME-PRESSURE ($\chi^2(1, 100) = 6.57$, $p = .010$, $OR = 0.34$, 95%$CI[0.15, 0.78]$).

This result is consistent with the idea that endowment reversals for undesirable options are only observed when the ability to engage in detailed comparisons between the choice options is limited. Taken together, the interaction between desirability and whether comparisons are encouraged or limited observed across Experiments 1, 2 and 3 provides strong support for the hypothesis that endowment reversals are more likely to occur when comparisons between the options are limited.

[3]As in the earlier experiments, we collapsed across which option was endowed when testing our key research questions. An exploratory analysis did not find a significant difference in **endowment preference** based on whether the 40% chance of $0.30 or 80% chance of $0.15 gamble was endowed in the TIME-PRESSURE conditions of Experiment 3 ($\chi^2(1, 100) = 0.50$, $p = .478$, $OR = 0.75$, 95%$CI[0.33, 1.68]$).

4.6 Experiment 4

The previous experiments established an important role of comparison processes in determining whether reversals of the endowment effect occur. These results are consistent with Dertwinkel-Kalt and Köhler's (2016) proposed dual-process theory, that combines mechanisms from salience theory and loss aversion. In the introduction of this chapter, we proposed an alternative theory that could explain these results based on a simple extension of prospect theory.

While we followed prospect theory in assuming that people evaluate options relative to a reference point, we proposed that this reference point is not only influenced by an individual's current state (e.g. possessing the endowed option) but also on an individual's previous reference state. If the previous reference state involved not possessing any related items (as was the case in our previous experiments), the resulting reference point would be more neutral than either the endowed or alternative option. Consequently, if comparisons are limited, participants should be simply deciding whether the endowed option is better than this more neutral reference point (leading to an endowment effect) or worse than this more neutral reference point (leading to an endowment reversal). On the other hand, if an individual has high expectations prior to a choice, the endowed option will be seen as inferior to the reference point, regardless of whether the subsequent choice options are desirable or undesirable, leading to endowment reversals. This hypothesis was tested in Experiment 4. It was therefore hypothesised that when participants were pre-exposed to an option that was more desirable than the subsequent choice options, they would show endowment reversals, favouring the non-endowed option regardless of the absolute desirability of the choice options. On the other hand, if a less desirable option was presented prior to the choice trial, participants would favour the endowed option regardless of the absolute desirability of the options.

4.6.1 Method

200 participants (50 per condition, mean age = 32 years old, 102 females) were recruited via Microworkers.com. The task was advertised as paying US$0.25 however, all participants were subsequently paid $0.75 for participation. Experiment 4 was a 2x2 between-subjects design crossing **desirability** (DESIRABLE or UNDESIRABLE) x **prior reference point** (SUPERIOR or INFERIOR reference point relative to the endowed gamble). Experiment 4 followed essentially the same design as the HIDDEN condition of Experiment 1 except for a few key changes. Firstly, the values of gambles shifted slightly. The endowed gamble was now a 50% chance of winning (losing) US$0.20 for all participants in the DESIRABLE (UNDESIRABLE) condition. The base rate that participants were told they were earning before partaking in the gamble was also shifted up slightly to $0.55. These changes were made to avoid the practice gamble approaching ceiling or floor levels in terms of percentage and

payout amounts while remaining within certain payout limitations implemented by Microworkers.com.

Secondly, and of key importance, Experiment 4 included an initial "practice" trial that participants took part in before they were then offered a choice of gambles. As in our previous experiments, participants were presented with two empty boxes, chose one and were presented with the supposed gamble corresponding to that box. However, participants did not choose whether they would want to keep that gamble. Instead, participants were now reminded that this choice was purely meant as practice; it was irrelevant to the subsequent choice and would not affect their payout in any way. The practice gamble was not resolved, meaning it had no effect on participants' payments and participants were not shown any outcome of the practice gamble. This practice trial was actually intended to induce a higher or lower reference point. Specifically, for half of the participants, this practice trial presented a gamble that was SUPERIOR to the endowed gamble in the following trial. In the DESIRABLE condition, this superior gamble was a 75% chance of winning \$0.30. In the UNDESIRABLE condition, this superior gamble was a 25% chance of losing \$0.10. For the other half of participants, this practice trial was INFERIOR to the subsequently endowed gamble. For DESIRABLE choices, the inferior gamble was a 25% chance of winning \$0.10 and for UNDESIRABLE choices the inferior gamble was a 75% chance of losing \$0.30.

To summarise, participants were first presented with a practice trial and were instructed that it was just an example and would not influence their payout. In this practice trial, participants were presented with a screen that was completely empty except for two boxes labelled "Gamble A" and "Gamble B" respectively. They were asked to choose one of the boxes before any information about the gambles was presented. Regardless of which box they chose, they were then presented with the practice gamble. They were not told the gamble corresponding to the other box. The practice trial then ended, without resolving the practice gamble, and the participants were invited to do the main experiment. As before, participants were presented with two boxes labelled "Gamble A" and "Gamble B" respectively and were invited to choose one of them. Regardless of which one they chose, they would be shown a gamble of 50% chance of winning (losing) \$0.20. They were not shown the gamble corresponding to the other box. Based on this information alone, they were then asked which gamble (i.e. A or B) they wished to take.

4.6.2 Results

Figure 4.1 on page 47 shows the observed **endowment preference** share and 95% confidence intervals for each of the conditions in Experiment 4. Were preferences towards the endowment influenced by the desirability of the options and the relative value of the practice gamble? A logistic regression was run predicting **endowment preference** (non-endowed gamble chosen = 0, endowed gamble chosen = 1) based on **desirability** (UNDESIRABLE = 0, DESIRABLE = 1) and **prior reference point** (INFERIOR = 0, SUPERIOR = 1).

As predicted, a significant main effect of the **prior reference point** was found, with the odds of choosing the endowed gamble being significantly higher when the practice gamble was INFERIOR than when it was SUPERIOR to the endowment ($OR = 5.25$, $p < .001$, $95\%CI[2.39, 13.46]$). We found no significant main effect of **desirability** ($OR = 1.34$, $p = .509$, $95\%CI[0.56, 3.23]$) and no significant interaction between **desirability** and **prior reference point** ($OR = 0.58$, $p = .365$, $95\%CI[0.17, 1.89]$). The data is consistent with our prediction that the participants' choices would be determined by the value of the gamble in the practice trial relative to the endowed gamble; participants kept the endowed gamble after seeing a practice gamble that was inferior to it but not when the practice gamble was superior.

4.7 Modelling the data

We present here a computational model based on prospect theory (Kahneman & Tversky, 1979). The goal is not to show that such a model offers the best explanation of our results, especially given we do not have sufficient data to verify a number of the model's more flexible assumptions. Instead, we wish to assess the extent to which, under reasonable assumptions, a model based on loss aversion can, quantitatively account for the existence of both the endowment effect and its reversal. This is especially important given previous claims that the endowment reversal is inconsistent with loss aversion (Brenner et al., 2007). The key addition to prospect theory in the model presented here is that we assume people's reference points are influenced not only by their current endowed state but also their previous reference state.

Prospect theory assumes that each option has a value associated with it representing the subjective evaluation of any potential outcomes weighted as a function of the probability of each outcome occurring (Kahneman & Tversky, 1979). Preference for one option over another is based on the difference in weighted value between these two options,where an option with a higher value will be preferred to one with a lower value (Kahneman & Tversky, 1979). To arrive at quantitative predictions of choice proportions we include noise, σ, as a free parameter in the model such that the final difference is logistically distributed with a mean equal to the difference in value between options and a standard deviation equal to the noise parameter (Wang & Fischbeck, 2004). Following Wang and Fischbeck (2004), Equation (4.1) shows the probability of preferring an endowed option, ϵ, over an alternative option, a, using a logistic function of the difference of the value of the two options:

$$P(\varepsilon|a) = \frac{1}{1 + e^{-\frac{V(\varepsilon)-V(a)}{\sigma}}} \tag{4.1}$$

Here, V is the overall subjective value of an option across its different possible outcomes. We define any option, x, as a set of outcomes, o^x, with a corresponding set of probabilities, p^x. V is defined as the sum of the subjective value v of each of the possible n outcomes of an option, weighted as a function π of the probability of each outcome occurring, as shown in Equation (4.2) (Kahneman & Tversky, 1979).

$$V(x) = \sum_{i=1}^{n} \pi(p_i^x) v(o_i^x | r) \tag{4.2}$$

Equation (4.2) acknowledges that the outcomes of x are evaluated relative to a reference point, r. The function $\pi(p_i^x)$ was proposed in prospect theory to weight the different possible outcomes of x by a nonlinear function of the probability of the outcome occurring (Kahneman & Tversky, 1979). For simplicity, we follow the lead of Kőszegi and Rabin (2006) and assume that people's evaluations of gambles is such that $\pi(p_i^x) = p_i^x$. This assumption is likely to breakdown for very small or very large probabilities, but should be reasonable for all other values (Kahneman & Tversky, 1979). In our gamble experiments, we avoided extremely large or extremely small probabilities for this reason. The function $v(o_i^x|r)$ calculates the subjective value of an outcome relative to the reference point, as outlined in Equation (4.3).

$$v(o_i^x | r) = \mu(o_i^x - r) \tag{4.3}$$

Equation (4.4) specifies the value function, μ. Specifically, as seen in Equation (4.4), we utilise the piecewise power function proposed by Tversky and Kahneman (1992), along with the median values they estimated for both the loss aversion parameter (2.25) and the exponent of the power function (0.88). These parameter estimates are used given their status as reasonable defaults to allow for model fitting in the absence of information to the contrary (e.g. Barberis & Xiong, 2009; Hens & Vlcek, 2011; Koop & Johnson, 2012).

This function satisfies the criteria for a value function set out by Kahneman and Tversky (1979) and Tversky and Kahneman (1991). In particular, it is a negatively accelerating function (due to the exponent of the power function being less than 1) for both gains and losses, but steeper in the losses domain than in the gains domain (due to the loss aversion parameter being greater than 1) to reflect loss aversion.

$$\mu(z) = \begin{cases} z^{0.88}, & \text{if } z \geq 0 \\ -2.25(-z)^{0.88}, & \text{if } z < 0 \end{cases} \tag{4.4}$$

To ensure our model was not reliant on the exact specification of the value function, we also fit the model with the value function used by Usher and McClelland (2004) in their Leaky Competing Accumulator model. Despite differences between the two value functions, they each provided essentially equivalent fits to our data (see Appendix B on page 153).

Equation (4.3) assumes that the reference point is certain but the option being considered can be risky. In the case that the reference point also has uncertainty, we must consider each of the possible outcomes of the choice option as well as each of the possible outcomes of the reference point in order to determine the value of the choice option relative to the reference point. Kőszegi and Rabin (2006, p. 1137) provide a general equation for such a case, which allows for integration across continuous probability distributions of different outcomes. Given we are only dealing with discrete probabilities in this study, and we have assumed that $\pi(p_i^x) = p_i^x$, we can simplify the equation used by Kőszegi and Rabin (2006) to represent the value of an option, x, with n possible outcomes relative to a risky reference point (r) with m possible outcomes to that shown in Equation (4.5).

$$V(x|r) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_i^x p_j^r v(o_i^x | o_j^r) \tag{4.5}$$

In prospect theory, the value of an option is determined relative to the reference point. A common assumption for choice tasks where one option is endowed is to assume that the reference point is equivalent to the endowed option, ε (Brenner et al., 2007; Koop & Johnson, 2012; Tversky & Kahneman, 1991). In this paper we argue that the reference point may not be determined entirely by the current endowed state. Instead, it may also be influenced by prior expectations.

In terms of how multiple pieces of reference information can be incorporated into the current reference state, Ordóñez, Connolly, and Coughlan (2000) found data consistent with the idea that people treat separate pieces of reference information as independent rather than averaging that information into a single reference point. Thus, a simple way in which the current reference state could incorporate information from both the previous reference state and the current endowed state is if people compare options to their current endowed state some of the time and compare options to the previous reference state the rest of the time. We can represent this with a discrete probability distribution where, at any point in time, the reference state r is set equal to the endowed option ε with probability γ, and is set equal to the previous reference state s otherwise. This is shown in Equation (4.6), where γ is a free parameter and $0 < \gamma < 1$.

$$\begin{aligned} P(r = \varepsilon) &= \gamma \\ P(r = s) &= 1 - \gamma \end{aligned} \tag{4.6}$$

Though Ordóñez et al. (2000) favoured the approach of having independent reference points, Appendix C on page 155 shows that, in our case, essentially equivalent fits can be obtained if we incorporate the previous reference state and current endowed state into a single reference point. Specifically, the model in Appendix C on page 155 assumes that the current reference state is represented as a probability distribution across possible outcomes, where beliefs about the probabilities of outcomes are updated based on previous expectations and the current endowment. Our goal is not to distinguish between such approaches but simply to show that loss aversion can explain the endowment reversals under various reasonable assumptions.

In Experiments 1, 2 and 3 participants did not have any basis for forming strong prior expectations of attaining non-zero outcomes. Thus, for these experiments, we set the previous reference state s as a 100% chance of winning \$0.00 (i.e. not expecting to win or lose any money) and, in the case of Experiment 2, expecting no change in the time required to complete the task either. In Experiment 4 participants viewed a practice gamble before completing the experimental choice in order to influence their previous reference state, s. Consequently, for Experiment 4, s in Equation (4.6) on the preceding page is set equal to the relevant practice gamble. For example, for the SUPERIOR DESIRABLE CONDITION of Experiment 4, s is set as a 75% chance of winning \$0.30 and a 25% chance of winning nothing.

We assume that participants treated the non-endowed alternative as equal to the reference point whenever the non-endowed option was hidden, or choices were made under time pressure. We made this assumption because the reference point represents the participant's expectations. If the non-endowed alternative was not processed, the participant had to decide whether or not to accept the endowed option based purely on whether it exceeded what they expected to get from switching.

To accommodate the fact that Experiment 2 used non-risky choice options, we firstly set $p_i^x = 1$ for these non-risky options given their outcomes were certain. Given the options in Experiment 2 could vary along two attributes (changes in the time commitment of the task and changes to monetary payout) we also included an additional free parameter, δ, to approximately equate these two attributes. Specifically, if an outcome i related to changes in the time commitment, the value presented to participants, $o_i^{x_{presented}}$, was multiplied by δ, as shown in Equation 7.

$$o_i^x = \begin{cases} \delta o_i^{x_{presented}}, & \text{when } o_i^{x_{presented}} \text{ relates to time commitment} \\ o_i^{x_{presented}}, & \text{when } o_i^{x_{presented}} \text{ relates to monetary payout} \end{cases} \tag{4.7}$$

4.7.1 Model fit

When fitting the model to our data, we set γ, σ and δ as free parameters and simultaneously fit the model to the endowment preference share across our experimental conditions. Values of the free parameters were found that maximised the likelihood of the data using the Nelder-Mead algorithm implemented by the "optim" function in R (R Core Team, 2016). The specific parameter values obtained were: $\gamma = 0.664$, $\sigma = 0.510$ and $\delta = 0.136$. The endowment preference share predicted by the model for each of the experimental conditions is presented in Figure 4.1 on page 47 and Table 4.1 on the facing page. A chi-square test failed to find a significant difference between the predicted endowment preference count based on the model and the observed endowment preference count, $\chi^2(9,\ N = 495) = 7.58,\ p = .870$. Our model can therefore adequately describe the data.

TABLE 4.1: Observed endowment preference share against predicted endowment preference share based on the presented model for each experiment in this chapter.

Experiment	Desirability	Comparisons	n	Endowment preference share	
				Observed	Model prediction
1	Desirable	Allowed	50	70%	69.9%
1	Undesirable	Allowed	50	76%	76.0%
1	Desirable	Hidden	50	58%	66.8%
1	Undesirable	Hidden	50	32%	33.2%
2	Desirable	Allowed	50	60%	55.8%
2	Undesirable	Allowed	50	56%	55.8%
2	Desirable	Hidden	50	72%	66.8%
2	Undesirable	Hidden	50	24%	33.2%
3	Desirable	Time-pressure	48	64.6%	66.8%
3	Undesirable	Time-pressure	47	38.3%	33.2%
4	Desirable	Inferior reference	50	60%	66.8%
4	Undesirable	Inferior reference	50	66%	66.8%
4	Desirable	Superior reference	50	32%	33.2%
4	Undesirable	Superior reference	50	26%	33.2%

4.8 Discussion

In previously published papers, the reversal of the endowment effect for undesirable options had only been observed for hypothetical choices (Brenner et al., 2007; Dertwinkel-Kalt & Köhler, 2016). To our knowledge, the current study represents the first time that such reversals have been observed for incentivised choices that have real consequences. However, we found that such reversals only occurred when comparisons between the endowed and non-endowed alternative were limited, whether that occurred by reducing the time available to compare the options or by hiding information about the non-endowed alternative. On the other hand, when comparisons were allowed and encouraged, participants generally preferred the endowed option regardless of desirability. We replicated these findings for risky (gambles to win or lose money) and non-risky (more money vs easier task) choices. We also showed that whether or not the endowment effect reversed depended on the participant's prior experience. In particular, when comparisons were limited, we showed that the endowment effect reversed for both desirable and undesirable gambles if the participant had previously been shown a more desirable practice gamble. Conversely, when shown a less desirable practice gamble, participants preferred the endowed gamble regardless of whether it was desirable or undesirable.

4.8.1 Theories of the endowment effect and its reversal

One of the theoretical points we wished to address in this paper was whether, as has been previously proposed (Brenner et al., 2007), endowment reversals are inconsistent with prospect theory and loss aversion more generally. We found that we were able to model all of the above findings using a straight-forward extension of prospect theory. While we cannot show whether the assumptions of this model are true, we were able to demonstrate that endowment reversals are, at least in principle, consistent with prospect theory.

In applying prospect theory to endowment reversals, we made the assumption that people have expectations that are influenced both by their current endowed state and their previous reference state. However, this explanation is not restricted to prospect theory and loss aversion. This assumption could also be incorporated into other theories of the endowment effect. For example, Gal (2006) suggests that people show inertia, in that they will not change their current state of affairs without sufficient motivation to do so. Gal proposes that it is this inertia, and the general difficulty in determining which option is best in order to motivate change, that leads to the endowment effect. If participants are comparing the endowed option to their more neutral expectations, it may be that an undesirable endowed option is sufficiently inferior to these more neutral expectations to motivate switching. While we have focused primarily on loss aversion in order to address previous claims that the reversed endowment effect is evidence against loss aversion, it is worth emphasising that such other mechanisms also provide viable explanations for our results.

We assumed above that participants' expectations (or reference points) were a function of their current endowed state and their previous reference state. However, it is worth considering the other reference points people could have held. A large number of these possible reference points, such as the expected value of the gambles, the best possible outcome or expecting an outcome of zero, are independent of the endowment and therefore do not help us explain the effect of endowing an option. Research into ambiguity aversion, the finding that people tend to favour known risks over unknown risks, might suggest that people are generally pessimistic, expecting outcomes that are worse than their known endowed state (Becker & Brownson, 1964; Camerer & Weber, 1992; Keren & Gerritsen, 1999). Counter to our results though, this would mean that participants will tend to prefer the known endowed option, regardless of desirability. Thus, our results cannot simply be explained by ambiguity aversion.

One alternative assumption about participants' expectations that could explain endowment reversals is that participants may perceive the non-endowed alternative to be riskier. For example, during the HIDDEN condition of Experiment 1 and the TIME-PRESSURE condition of Experiment 3, participants may have expected the non-endowed gamble to have approximately the same expected value as the endowed gamble but expected that the non-endowed gamble could just as likely have small potential outcomes as extreme potential outcomes. Importantly, there is evidence that people are risk-seeking for losses and risk-averse for gains (Kahneman & Tversky, 1979). This means that risk seeking during undesirable scenarios could explain preference for the non-endowed gamble while aversion to this risk might explain the observed preference for the endowment in the DESIRABLE condition. Thus, it is possible that perception of the non-endowed option as riskier could explain, or have contributed to, the endowment reversals we observed in Experiments 1, 2 and 3. Nevertheless, additional mechanisms would still be needed to explain why the endowment effect occurs when comparisons are encouraged and why a practice gamble can affect whether or not the endowment effect reverses, as observed in Experiment 4.

We have focused thus far on how these results can be explained by loss aversion given loss aversion represents the most prominent explanation of the endowment effect. However, other mechanisms may be underlying these results. In Chapter 2 we discussed a number of other theories that could explain the endowment effect. For example, pre-decisional information distortion suggests that decision makers tend to interpret available information such that it favours the endowment (Willemsen et al., 2011; Carmon & Ariely, 2000). This may generally be the case except for when the endowment is worse than a decision maker's reference point and is not comparing the endowed option directly to an alternative. In such a case, the fact that the endowment is worse than expectations may cause information to be distorted against the endowment. Thus, our results do not necessitate loss aversion but instead emphasise the need to account for how people are processing the available options and the important impact that people's expectations have on their response to endowments.

With that said, the theories our results do argue against are those that assume endowing an item results in perceptions of that item being amplified. Brenner et al.'s (2007) possession loss aversion and attribute weighting models like salience theory (Bordalo et al., 2012a) or the associative accumulation model (Bhatia, 2013) assume that the value of an endowed item is amplified such that, if it is undesirable, it will be avoided in favour of an alternative undesirable item. The fact that we found no endowment reversals when participants were allowed to and encouraged to freely compare the available choice options cannot be explained by these theories.

4.8.2 Implications

The occurrence of endowment reversals shows that endowing an item can have completely opposite effects on preference for that item; potentially either increasing or decreasing its perceived value. Understanding when these endowment reversals are likely to occur is particularly important whenever trying to use ownership as a way of manipulating behaviour, either in laboratory studies or as part of real-world nudging strategies. We have shown that these endowment effect reversals can occur for choices with real, albeit small, consequences. Furthermore, they can occur for both desirable and undesirable options, dependent on the decision maker's expectations.

The fact that we only observed endowment reversals when comparisons were limited also has interesting implications for how we study endowments. For example, the mode in which choice information is presented could be expected to influence the extent to which options are compared and the subsequent likelihood of endowment reversals occurring. Choice information presented in tables tends to elicit detailed comparisons (Noguchi & Stewart, 2014) while information presented in blocks of text is difficult to extract and compare (Kelly, 1993; Smerecnik et al., 2010). Thus, presenting choices in text may increase the potential for endowment reversals to occur compared to when options are presented in a table.

This may explain why we failed to observe an endowment reversal (or reversals of any other choice biases) in Chapter 3. Though we presented participants with hypothetical choices, just as Dertwinkel-Kalt and Köhler (2016) and Brenner et al. (2007) did, our choice options were presented in a table that encouraged participants to directly compare the options. In contrast, Dertwinkel-Kalt and Köhler (2016) and Brenner et al. (2007) described the available options in paragraphs of text, making it harder for participants to compare the options. Making it easy to compare options in Chapter 3 may have made it easier for participants to engage in the choice even without explicit incentives.

Individuals may also systematically differ in their tendency to show endowment reversals. People with high need for cognition, for example, tend to engage in deeper and broader comparisons in decision situations (Levin, Huneke, & Jasper, 2000). Thus, low need for cognition may be associated with a greater tendency to show endowment reversals due to a tendency not to compare the options. Individual differences can be crucial to our understanding of how manipulations are affecting choice (e.g. Liew et al., 2016). Need for cognition offers a potentially interesting individual-level variable that could explain variation in how endowments affect choices for different people

4.8.3 Conclusion

In this study, we set out to provide a greater understanding of when the endowment effect is likely to reverse and, thus, the theoretical implications of these reversals. Two factors were found to influence whether endowments increased or decreased preference: how the choice options were being compared and the likely expectations held when going into the choice. Crucially, we showed that these conditions in which endowment reversals occur can actually be accounted for by theories that were previously considered inconsistent with endowment reversals (e.g. prospect theory). These results emphasise the importance of considering how people are processing information and their expectations both when studying the endowment effect and if attempting to implement endowment manipulations in real-world nudging strategies. Simple changes in the extent to which options are compared and the expectations people carry into a decision situation can determine whether ownership increases or decreases preference for an option.

Chapter 5

Understanding Norm Effects: A Preface to Chapter 6

In the previous chapter, we aimed to provide some clarity on findings relating to the mechanisms underlying the endowment effect. Much of the interest in the endowment effect relates to the fact it tells us something interesting about how people make choices; favouring an option more when it is endowed violates normative models of choice and thus, the existence of the endowment effect suggests that other mechanisms influence our preferences. Accordingly, there has been a notable amount of research and interest in the mechanisms driving the endowment effect. In contrast, the choice bias discussed in the upcoming chapters (the descriptive norm effect) has, comparatively, received almost no theory-driven testing.

Social norms can be broken down into a number of different forms. The two major forms of normative influence are how others believe we should behave (injunctive norms) and how other people actually behave (descriptive norms; Cialdini et al., 2006). We focus on descriptive norms in the upcoming chapters, primarily because they are the easiest to manipulate in an experimental setting; it is much less contentious for me to say that 75% of people chose an option than to try to convince you that 75% of people believe you should choose an option. For example, Stok, De Ridder, De Vet, and De Wit (2014) found evidence that teenagers reacted negatively to injunctive norms favouring fruit consumption, showing lower intentions to consume fruit than those exposed to descriptive norms or even no norm at all. Nevertheless, Eriksson et al. (2015) found, across multiple paradigms, that people tend to infer the existence of injunctive norms when informed about descriptive norms and vice versa. Thus, without going out of our way to distinguish these two types of norms, it is quite possible that manipulating descriptive norms unintentionally also influences people's injunctive normative beliefs.

It comes as a surprise to nobody that social norms tend to influence people's behaviour. Nolan, Schultz, Cialdini, Goldstein, and Griskevicius (2008) found that beliefs about descriptive norms had a stronger effect on energy conservation behaviour than a range of other beliefs (e.g. saving money, environmental protection and social responsibility), despite people rating normative beliefs as the least motivating factor. Despite the prevalence of social norms in society, theory-driven testing of social norm effects is scarce. Instead, an extensive literature has developed looking at when and where these effects can influence importance decisions (Alm et al., 2019; Ecker et al., 2019; Passafaro et al., 2019; Vinnell et al., 2019). These applied studies have established the important role that descriptive norms can have on people's preferences but offer little insight into the mechanisms driving the effect.

This strong focus on applied study of the descriptive and injunctive norm effects instead of theory-driven testing may largely be a consequence of the fact that these biases have some straightforward explanations. A standard explanation for the descriptive norm effect is that the popularity of an option implies that other people have determined that that option is better than alternatives (Deutsch & Gerard, 1955; Cialdini et al., 1990). Thus, according to this informational account, the popularity of an option provides us with useful information as to which option is best.

While this informational account suggests that descriptive norms provide insight into the value of different options, the social sanction account proposes that norms themselves inherently affects the value of options. Specifically, if most other people favour or approve of a particular option, they may potentially respond positively to you if you choose the same option and negatively to you if you instead violate the norm. Thus, even if another option might be otherwise preferable to you, the existence of a norm towards a certain behaviour can offer an objective reason to conform to the norm so as to avoid negative social sanctions.

The fact that norm effects can be explained through norms offering objective reasons why people should conform to them, means that they are potentially less informative about the human decision making process than other choice biases that strictly violate axioms of rational choice theories. For example, imagine you are offered a choice between a blue and red box, one of which contains $1,000 and the other contains $1. You do not get to look inside the boxes but you are informed that, when offered the same choice, other people all chose the blue box. With this knowledge, you can infer that other people may have known that the blue box contained the larger sum and thus, you should choose the blue box too. While this is nice for you in that the norm helps you win $1,000, it tells us essentially nothing about your decision making process other than the fact that you are capable of some basic theory of mind to determine that other people chose the best box and so you should too. Thus, it may well be due to these objective reasons that norm effects have received relatively little theory-driven testing.

While it is easy to come up with these trivial examples where people should obviously conform to social norms, this does not necessarily mean that these objective explanations fully account for the descriptive norm effect. In Chapter 6, we look at the effect of descriptive norms that have arisen for random, arbitrary reasons, offering no useful information and carrying no threat of social sanctions. Despite there being no objective informational or social sanction based reason to follow these arbitrary norms, we find that participants still tend to conform to them. This finding suggests that the descriptive norm effect is not strictly the result of norms offering objective reasons why we should conform to them. Instead these results indicate that the descriptive norm effect is influenced by other mechanisms that are engaged even when social norms are arbitrary. Chapter 8 looks at whether self-categorisation theory, the most prominent explanation for the descriptive norm effect outside of the informational and social sanction accounts, can explain this effect.

Chapter 6

Even Arbitrary Norms Influence Moral Decision-Making

The contents of this chapter are adapted from our paper published in *Nature Human Behaviour*: Pryor, C. G., Perfors, A., & Howe, P. D. L. (2019b). Even arbitrary norms influence moral decision-making. *Nature Human Behaviour*, *3*, 57–62

6.1 Abstract

This chapter looks at the theoretical mechanisms underpinning people's tendency to conform to behaviour that is popular amongst other people (descriptive norms). Here we show that people conform to descriptive norms, even when they understand that the norms in question are arbitrary and do not reflect the actual preferences of other people. These results hold across multiple contexts and when controlling for confounds such as anchoring or mere exposure effects. Two prominent explanations of norm adherence, the informational and social sanction accounts, cannot explain this observed effect of arbitrary descriptive norms. Moreover, we demonstrate that the degree to which participants conform to an arbitrary norm is influenced by the degree to which they self-identify with the group that exhibits the norm. We discuss how these results can be explained by self-categorisation theory.

6.2 Introduction

It is well-known that individuals tend to copy behaviours that are common amongst other people, a phenomenon known as the descriptive norm effect (Cialdini et al., 1990). For example, Burger et al. (2010) found that participants were more likely to choose a healthier snack when shown that previous participants chose healthy snacks rather than candy bars. Applied studies have found that informing people of the relevant descriptive norms can increase attendance at doctor appointments (Hallsworth et al., 2015), decrease tax evasion (Wenzel, 2005), increase towel reuse at hotels (Goldstein et al., 2008), decrease long-term energy use (Schultz et al., 2007) and increase organ donor registrations (Behavioural Insights Team, 2013). However, despite widespread interest in the descriptive norm effect and its use for encouraging real-world, pro-social decisions, it is still unclear why we conform to descriptive norms.

Two of the most prominent explanations of the descriptive norm effect focus on two objective reasons for following descriptive norms, first delineated by Deutsch and Gerard (1955). According to the informational account, people might choose to conform to the descriptive norm because other people's choices are informative about what is likely to be "effective and adaptive action" (Cialdini et al., 1990, p. 1015). If everyone else has chosen to behave a certain way, it is likely because they have determined that that behaviour offers some greater benefits than alternative behaviours. This knowledge alone therefore offers objectively rational reasons to conform to the descriptive norm, so that you too can access these benefits.

The social sanction account offers another objective reason to conform to descriptive norms. Specifically, the social sanction account suggests that when people prefer a particular behaviour they may respond negatively to you if you fail to act in line with this preference. People might therefore conform to descriptive norms in order to reduce the possibility of negative social sanctions (Ajzen & Fishbein, 1980; Bendor & Swistak, 2001).

Both of these accounts of the descriptive norm effect assume that other people's preferences are relevant to how an individual should objectively act. In line with this, previous research on the descriptive norm effect has focused on descriptive norms that arise through people's choices. However, there are often situations in which a lack of available alternatives or a lack of incentive to explore these alternatives can lead to a descriptive norm occurring. For example, when a default option is available, people may adopt the default simply because they do not care enough about the decision to act differently, such as subscribing to a company's default pension plan (Choi, Laibson, Madrian, & Metrick, 2005). This raises the question of whether people will conform to norms that have arisen arbitrarily and do not reflect other people's actual preferences. Crucially, from an objective perspective, people would not be expected to follow obviously arbitrary descriptive norms that they know do not reflect people's preferences because such norms provide no useful information about the value of the options or about potential social sanctions.

Self-categorisation theory provides a very different explanation of the descriptive norm effect (Turner et al., 1987). This theory proposes that people's self-identity is often heavily influenced by the social groups that the individual identifies with (their 'ingroups'). In an effort to maintain this sense of group identity, individuals are expected to engage in "behaviours that optimally minimise ingroup differences and maximise intergroup differences" (Terry & Hogg, 1996, p. 779). Consequently, if their ingroup displays a prominent behaviour that is different from the behaviour displayed by the outgroup, self-categorisation theory predicts that the individual will tend to conform to this behaviour to reinforce their identity as a member of the ingroup.

Crucially, there is nothing in self-categorisation theory that requires the behaviour to reflect the actual preferences of the ingroup. Rather, self-categorisation theory predicts that conformity will occur whenever the behaviour is salient and characteristic of the ingroup (Turner et al., 1987). As explained by Terry and Hogg (1996, p. 780), individuals will conform to the norms of a particular group whenever membership of that group is "the contextual basis for [their] self-definition". Thus, self-categorisation theory predicts that an individual should conform to salient descriptive norms, even if they are arbitrary and do not reflect other people's actual preferences, providing they are characteristic of a salient ingroup.

6.2.1 Aims and hypotheses

The aim of this study is to distinguish between the prominent objectivity-based accounts and the self-categorisation account of the descriptive norm effect by testing whether people conform to arbitrary descriptive norms. We present descriptive norms of what previous participants allegedly did. However, we explicitly state that these previous participants had not chosen their course of action but were instead randomly allocated to certain behaviours and asked to reason about what they thought of them. By emphasising that these previous participants did not choose to behave in the manner in which they did, we make it clear that these norms did not reflect their beliefs, preferences or values and thus had no informational or social sanction implications. Consequently, of the prominent explanations of the descriptive norm effect, only self-categorisation theory would predict an effect of these arbitrary descriptive norms.

To preempt our results, we found that participants did conform to arbitrary descriptive norms even when controlling for a number of possible confounds. This suggests that the information and social sanction accounts cannot provide a complete description of the descriptive norm effect and need to be supplemented by an additional mechanism. To better understand the nature of this additional mechanism, our final experiment investigated whether individuals who more strongly identified with a particular group conformed more to the arbitrary descriptive norms of that group. Consistent with self-categorisation theory we found that they did. This is studied further in Chapter 8.

6.3 Experiment 1

This experiment investigated whether participants would conform to arbitrary descriptive norms that do not reflect the actual preferences of previous participants. It was run using Amazon's Mechanical Turk (MTurk) because MTurk participants tend to cover a wider range of ages and ethnicities than the traditional university participant pool (Behrend, Sharek, Meade, & Wiebe, 2011) and match the general population closely in terms of the distribution of occupations; and although they are slightly younger than the general population, they are much closer to that standard than university undergraduates (Huff & Tingley, 2015). From a quality of data point of view, Mechanical Turk is similar to that of a traditional university participant pool. Hauser and Schwarz (2016) found that MTurk participants actually outperformed laboratory participants in their comprehension of instructions. Findings from laboratory studies tend to replicate in MTurk samples (Horton, Rand, & Zeckhauser, 2011; Huh, Vosgerau, & Morewedge, 2014; Paolacci, Chandler, & Ipeirotis, 2010) and the test-retest reliability of MTurk data is generally quite high (Buhrmester, Kwang, & Gosling, 2011). Kittur, Chi, and Suh (2008) reported that failing to include understanding checks can substantially reduce the quality of the data, though their effect size may have been inflated by their task being particularly conducive to bots and spammers. Regardless, understanding checks were included in all of our experiments.

6.3.1 Method

Participants

150 English-speaking participants (mean age = 36 years old, 43% female) were recruited via Amazon's Mechanical Turk and provided informed consent to participate. Participants completed the experiment in a web browser and were paid US$0.65 for participating. The experiment took around 2-3 minutes to complete.

We conducted a power analysis using 10,000 bootstrapped samples from a pilot study we ran on 105 participants, which was largely the same as Experiment 1. Power was calculated as the proportion of bootstrapped samples that had a 95%CI excluding OR = 1 using the ordinal logistic regression outlined below. This power analysis suggested that a sample size of 72 would provide over 80% power to observe a 95%CI that excludes OR = 1. The (unreported) pilot study did not include an understanding check but, as outlined below, all of the experiments we ran included a pre-planned understanding check as an exclusion criterion. To ensure adequate power even in the face of losing some participants who fail the understanding check, we increased the sample size to 150 for each of our experiments.

Procedure

Upon agreeing to participate, participants provided some demographic information and completed a 10-item personality questionnaire taken from Gosling, Rentfrow, and Swann (2003). Based on their responses, participants were presented with a "personality profile" that informed them of which two Big Five personality traits (openness, agreeableness, emotional stability, extraversion and conscientiousness) they scored highest on.

As background to justify the arbitrary descriptive norms presented later in the experiment, participants were then informed that we were following up on a previous paper that looked at how people feel during moral dilemmas. Crucially, they were told that participants in the previous study were not asked to choose how they would act in the moral dilemma but instead were randomly allocated to imagine performing a specific action and to then rate how good or bad they would feel about performing that action. After reading the instructions, participants proceeded to the experimental trial, where they were presented with the following moral dilemma:

> Imagine you have witnessed a man rob a bank. However, you then saw him do something unexpected with the money. He donated it all to a run-down orphanage that would benefit greatly from the money. You must decide whether to call the police and report the robber or do nothing and leave the robber alone.

Immediately below this moral dilemma was a normative statement. Importantly, this statement made it clear that the norm had not arisen through choices but instead through an error-prone random allocation process, which meant that it provided no information about the actual beliefs or values of the previous participants. The specific wording of the normative statement was:

> In the previous study, approximately 75% of [GENDER] participants aged [AGE-RANGE] that rated high in [PERSONALITY DESCRIPTION] were allocated to [RANDOMLY-POPULAR ACTION] and then rate how they would feel, due to an error in their random allocation.

When presented to participants, [GENDER] was replaced with the participant's gender, [AGE-RANGE] was replaced with the participant's five-year age-range (e.g. 30-35) and [PERSONALITY DESCRIPTION] was replaced with the two personality traits that the participant had rated high on (e.g. extraversion and agreeableness). Also, [RANDOMLY-POPULAR ACTION] was replaced with one of the available actions, varied randomly between participants. Thus, half of the participants were informed that the [RANDOMLY-POPULAR ACTION] was to "call the police and report the robber" (the REPORT condition) while the remaining half were informed that it was to "do nothing and leave the robber alone" (the NOT REPORT condition).

Participants were then asked about what action they would choose by responding on the following 6-point Likert scale:

- Definitely call the police and report the robber
- Very likely call the police and report the robber
- Probably call the police and report the robber
- Probably do nothing and leave the robber alone
- Very likely do nothing and leave the robber alone
- Definitely do nothing and leave the robber alone

After completing this main experimental trial, participants rated how good or bad they would feel about performing their chosen action. This was included purely for the sake of consistency with the backstory offered in the introduction and was not analysed. Finally, participants completed an understanding check, asking them to identify the problem with the previous study that we described in the instructions. This was included to ensure that participants believed that the norm they saw really was arbitrary and did not mistakenly think that it reflected the previous participants' preferences. The response options in the understanding check, presented in a random order, were:

1. An error meant that more participants were allocated to rate one action than the other (correct)
2. One action was preferred by most participants, while very few participants said they were likely to perform the other action (incorrect)
3. No data was saved during the experiment (incorrect)
4. The participants completed the experiment with their eyes closed (incorrect)

Design and Analysis

A one-shot between-subjects design was used. The independent variable was the randomly-allocated descriptive norm; whether more people had been randomly allocated to rate reporting the robber (REPORT condition; n = 75) or not reporting the robber (NOT REPORT condition; n = 75). The dependent variable was participants' responses on the Likert scale rating the certainty with which they would act a certain way.

A two-tailed ordinal logistic regression was run to assess the extent to which preferences towards reporting the robber increased for participants in the REPORT condition compared to those in the NOT REPORT condition. It is perhaps worth noting that a common approach to analysing Likert scale data is to treat it as interval data and analyse it with a general linear model. However, it is quite often incorrect to assume that there are equal intervals between ordered categories and that residuals will even be approximately normally distributed. The problems with treating ordinal data as continuous is well outlined by Liddell and Kruschke (2018), who found that analysing simulated ordinal data with general linear models lead to inaccurate effect size estimates and greatly increased false alarm rates. We therefore analyse the data using an ordinal logistic regression, which makes simpler, more defensible assumptions and should mitigate the risk of inflated false alarm rates or misleading effect sizes.

Obviously, it would not be interesting if people conform to the norm only because they were under the misapprehension that it reflected the true beliefs and values of people like themselves. Consequently, we excluded from the analysis any participant that answered the single-question understanding check incorrectly; however, as we describe below, our conclusions remain the same when all participants are included or even when we only exclude participants that chose option #2 above.

Data Availability

The response data for all experiments in this paper are available in the following OSF repository: https://osf.io/n6uz5/?view_only=7dc67fcc0c1f4fdea8fe1dfe5d492480

6.3.2 Results

Does being presented with a norm that has arisen simply through random allocation shift preferences in line with the norm? Figure 6.1 on the next page addresses this question by showing the proportion of responses in each condition. An ordinal logistic regression[1] run on participants who answered the understanding check correctly found a significant effect of the randomly-allocated norm on responses ($N = 74$, $OR = 3.64$, $p = .003$, $95\%CI[1.58, 8.65]$): that is, people shifted their own responses toward the descriptive norm provided by the previous participants, even though they knew this descriptive norm was arbitrary. The observed odds ratio shows that the odds of choosing an option more in line with reporting the robber were approximately four times higher for participants in our REPORT condition than those in our NOT REPORT condition. These results remain qualitatively the same if we include all participants regardless of their response on the understanding check ($N = 150$, $OR = 2.16$, $p = .009$, $95\%CI[1.22, 3.85]$) or if we only exclude those participants who chose the understanding check option corresponding to believing that the norm reflected other participants' choices ($N = 116$, $OR = 2.30$, $p = .013$, $95\%CI[1.20, 4.46]$.

6.4 Experiment 2

The purpose of this experiment was to investigate whether the results of Experiment 1 would replicate when a different moral dilemma was used. The moral dilemma in Experiment 1 involved parties that were unknown to the decision maker, potentially reducing the engagement of participants with the dilemma due to a lack of personal involvement. For example, Greene, Sommerville, Nystrom, Darley, and Cohen (2001) found that brain areas associated with emotion were more strongly activated by personal than impersonal moral dilemmas. Thus, we looked at whether our results from Experiment 1 replicated using a moral dilemma in which the decision maker had a personal relationship with the affected party.

[1] In this, and all subsequent experiments reported in this paper, we failed to find sufficient evidence ($p > .05$) to reject the proportional odds assumption of the ordinal logistic regression (that the effect of the norm was the same across response thresholds). Similarly, goodness-of-fit tests failed to find significant misfit ($p > .05$) between the predictions of the ordinal logistic regression model and the data in any of our experiments.

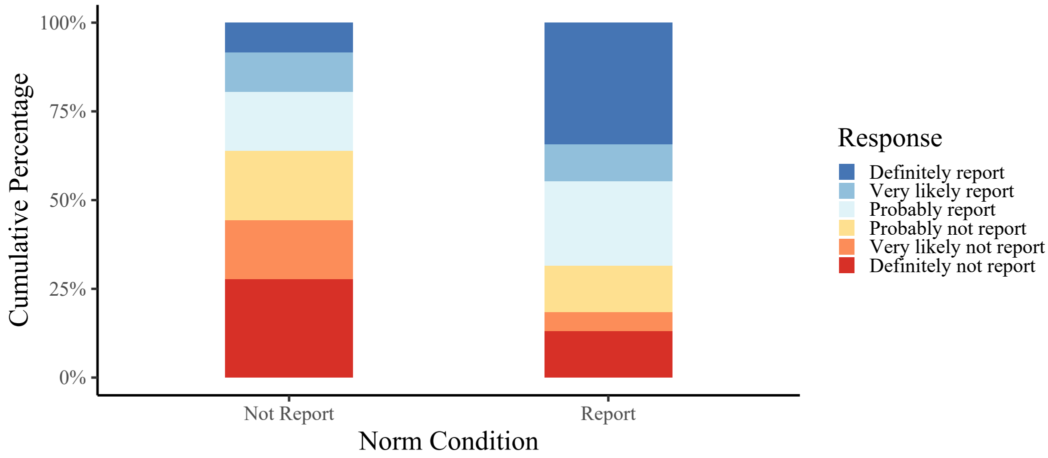

FIGURE 6.1: Proportion of responses in Experiment 1 for both the REPORT and the NOT REPORT norm conditions. Participants shifted their responses in the direction of the norm they were given, even though they knew the norm was arbitrary.

6.4.1 Method

Participants

151 English-speaking participants (mean age = 35 years old, 40% female) who did not participate in Experiment 1 were recruited via Amazon's Mechanical Turk and provided informed consent to participate. Participants completed the experiment in a web browser and were paid US$0.65 for participating.

Procedure and Design

Experiment 2 was exactly the same as Experiment 1 except participants were now presented with a different moral dilemma:

"Imagine you are hiring someone for a job at your firm. Your friend has applied for the position and is qualified. However, another applicant seems to be even more qualified."

This time participants were presented with a norm that either stated that more people had been allocated to imagine "hiring the more qualified candidate" ($n = 76$) or "hiring their friend" ($n = 75$) and the response options were updated accordingly.

6.4.2 Results

The proportion of participants in each condition that chose each response option is shown in Figure 6.2 on the facing page. As in Experiment 1, after excluding any participant who failed the comprehension check, participants shifted their responses towards the direction of the arbitrary descriptive norms provided by previous participants; an ordinal logistic regression found that the odds of showing stronger preference towards hiring the more qualified candidate were significantly higher for participants who were told that most previous participants were randomly allocated to hiring the more qualified candidate, and vice-versa ($N = 93$, $OR = 2.84$, $p = .006$, $95\%CI[1.35, 6.08]$). As before, we obtained qualitatively the same result if we included all participants ($N = 151$, $OR = 1.98$, $p = .019$, $95\%CI[1.12, 3.53]$) or excluded only those participants who indicated belief that the norm arose through participant choices ($N = 105$, $OR = 2.88$, $p = .003$, $95\%CI[1.43, 5.93]$).

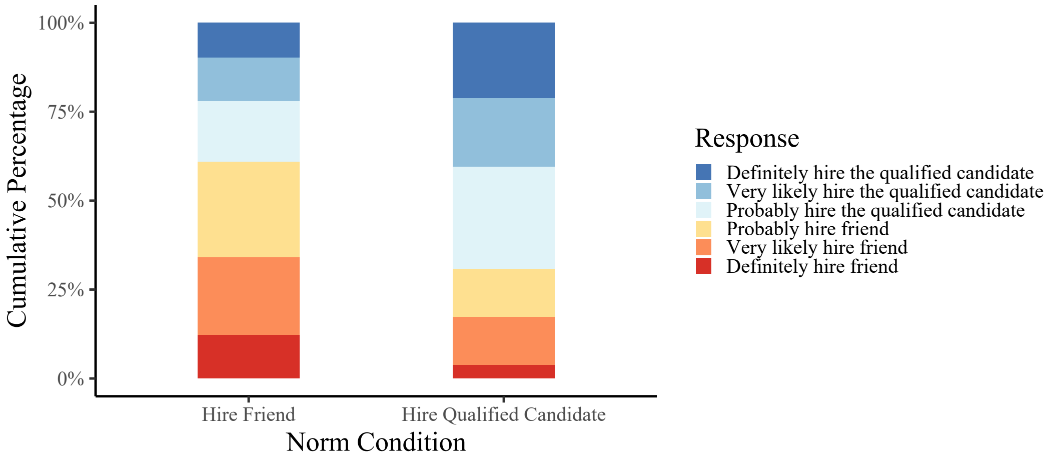

FIGURE 6.2: Proportion of responses in Experiment 2 for both the HIRE FRIEND and the HIRE QUALIFIED CANDIDATE norm conditions. As in Experiment 1, participants' responses shifted in the direction of the norm they were given, even though they knew the norm was arbitrary.

6.5 Experiment 3

While our results so far suggest that people conform to descriptive norms even when they do not reflect the actual beliefs and values of the ingroup, it is possible that these results arose simply because the normative statement referred to which option had arbitrarily become "popular" (the RANDOMLY-POPULAR option) rather than which option was correspondingly less popular (the RANDOMLY-UNPOPULAR option). The mere-exposure effect means that simply increasing exposure to an item can increase preference for that item (Zajonc, 1968). Thus, it may be that the increased preference for the RANDOMLY-POPULAR option arose simply due to its increased exposure.

Experiment 3 tested this prediction by stating that an action had "approximately 25%" of similar participants allocated to it. If participants were previously conforming to the norm due to mere-exposure effects, then we would expect them to now show a preference for this randomly-unpopular action, given that is now the option they are exposed to. On the other hand, if they are actually conforming to the norm, then they should still prefer the randomly-popular action (i.e. the action not described in the normative statement this time).

6.5.1 Method

Participants

150 English-speaking participants (mean age = 36 years old, 54% female) who did not participate in Experiments 1 or 2 were recruited via Amazon's Mechanical Turk and provided informed consent to participate. Participants completed the experiment in a web browser and were paid US$0.65 for participating.

Procedure and Design

Experiment 3 was exactly the same as Experiment 1 except that that the phrase "approximately 75%" in the normative statement was replaced with "approximately 25%". For simplicity, we will continue to refer to variables in terms of which action is RANDOMLY-POPULAR ($n = 75$ per condition). For example, participants in the REPORT condition were shown a normative statement that the option to "do nothing and leave the robber alone" had "approximately 25%" of similar participants allocated to it, implying that most similar participants were allocated to imagine and rate reporting the robber.

6.5.2 Results

Do people conform to arbitrary, randomly-allocated descriptive norms even when the normative statement relates to the RANDOMLY-UNPOPULAR option? Excluding all participants who failed the comprehension check, as Figure 6.3 on the next page shows, participants did indeed still conform to the arbitrary descriptive norm even when it was expressed in terms of which option was less common. An ordinal logistic regression found that the odds of choosing an option that more strongly favours reporting the robber were significantly higher for participants in the REPORT condition (who were told that relatively few people were allocated to not report the robber) than those in the NOT REPORT condition ($N = 72$, $OR = 4.68$, $p < .001$, $95\%CI[1.95, 11.72]$). As before, we draw the same conclusions if we include all participants regardless of their response on the understanding check ($N = 150$, $OR = 2.08$, $p = .0127$, $95\%CI[1.17, 3.72]$) or if we exclude only those participants who chose the understanding check option corresponding to believing that the norm arose through participant choices ($N = 122$, $OR = 2.39$, $p = .008$, $95\%CI[1.26, 4.59]$). These results are not consistent with a mere-exposure explanation, instead suggesting that people like to do what others are doing and are not simply choosing the option that they have been exposed to more.

6.6 Experiment 4

Experiment 4 tested whether our results could be explained by participants systematically misreading the normative statement. The fact that we found descriptive

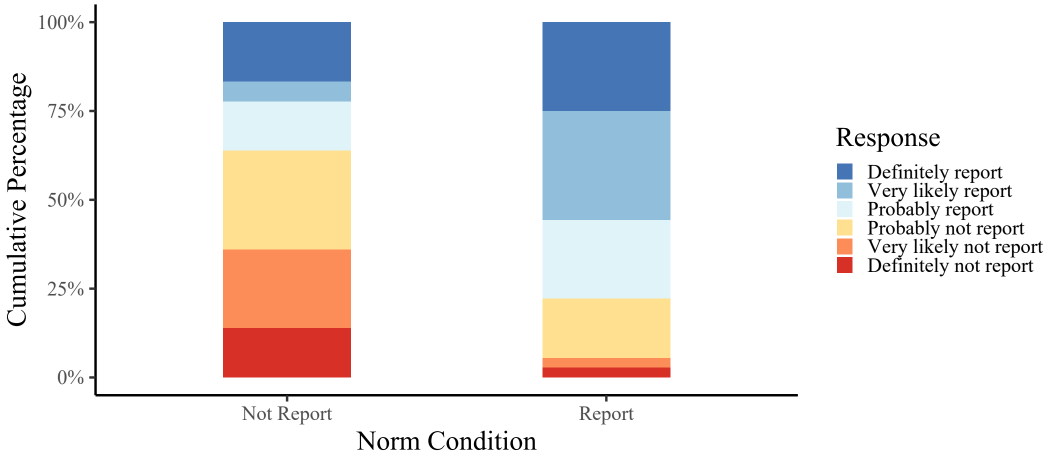

FIGURE 6.3: Proportion of responses in Experiment 3 for both the REPORT and the NOT REPORT norm conditions. As in Experiments 1 and 2, participants' responses shifted in the direction of the norm they were given, even though they knew the norm was arbitrary.

norm effects even when we excluded participants who failed the understanding check provides evidence against such an explanation. However, it is still possible that participants focused primarily on the percentage information (e.g. "75%") and the corresponding action (e.g. "call the police and report the robber"), skipping over the rest of the normative statement during the trial. Thus, at the time of making a choice, they may have implicitly treated the norm as reflecting preferences, ignoring the fact that it was random.

Experiment 4 tested this explanation by altering the normative statement such that it no longer related to the current moral dilemma. If participants are systematically misreading the normative statement, then they should not be sensitive to such changes and we would expect them to continue to conform to this now-unrelated descriptive norm. On the other hand, if participants fully comprehend the normative statement, then they should recognise that it does not relate to the current moral dilemma and ignore it when making a decision.

6.6.1 Method

Participants

150 English-speaking participants (mean age = 35 years old, 44% female, $n = 75$ per condition) who did not participate in the previous experiments were recruited via Amazon's Mechanical Turk and agreed to participate in exchange for US$0.65.

Procedure and Design

Experiment 4 was exactly the same as Experiment 1 except that the normative statement now informed participants that the supposed previous study presented participants with an unrelated moral dilemma. The specific wording of the normative statement was as follows:

> The previous study used a completely different moral dilemma that happened to also involve a robber. In the previous study, approximately 75% of [GENDER] participants aged [AGE-RANGE] that rated high in [PERSONALITY DESCRIPTION] were allocated to [RANDOMLY-POPULAR ACTION] and then rate how they would feel, due to an error in their random allocation.

6.6.2 Results

The proportion of responses in each condition are shown in Figure 6.4. Were participants mistakenly influenced by normative statements that did not relate to the current choice? After excluding any participant who failed the comprehension check, participants did not prefer the RANDOMLY-POPULAR action when the norm was derived from an unrelated moral dilemma ($N = 91, OR = 0.79, p = .524, 95\%CI[0.38, 1.63]$). We draw the same conclusions if we include all participants regardless of their response on the understanding check ($N = 150, OR = 0.83, p = .512, 95\%CI[0.47, 1.46]$) or if we only exclude those participants who chose the understanding check option corresponding to believing that the norm arose through participant choices ($N = 109, OR = 0.75, p = .395, 95\%CI[0.38, 1.45]$).

In fact, this effect is significantly weaker than the effect of the norm observed in Experiment 1, as evidenced by an interaction between which action was RANDOMLY-POPULAR and the experiment ($N = 165$, $OR = 0.21$, $p = .005$, $95\%CI[0.07, 0.62]$). Given the only difference between these experiments was that Experiment 4 stated that the norm was not related to the current moral dilemma, this suggests that the conformity to arbitrary descriptive norms observed in previous experiments was not simply the result of failing to read the norm statement properly.

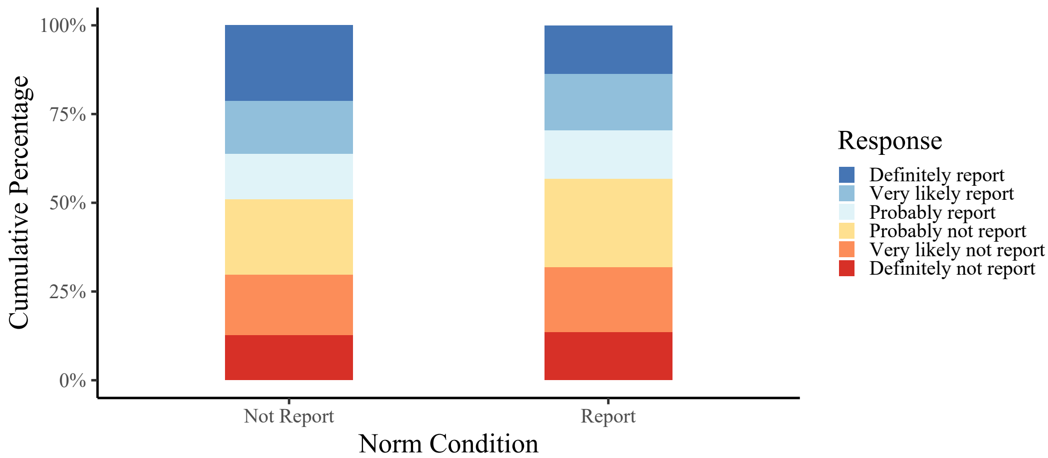

FIGURE 6.4: Proportion of responses in Experiment 4 for both the REPORT and the NOT REPORT norm conditions. There was no significant effect of the unrelated arbitrary descriptive norm on participants' responses. This suggests that the results from the previous experiments did not simply reflect a failure to read the norm statement properly.

6.7 Experiment 5

The previous experiments found that people conform to arbitrary descriptive norms, suggesting that a mechanism that does not require the norms to reflect actual preferences must contribute to the descriptive norm effect. An open question then is what such a mechanism might be. Self-categorisation theory offers one potential mechanisms through which people are driven to follow descriptive norms even when they are arbitrary. Self-categorisation emphasises the importance of identifying with a group when conforming to their norms. Experiment 5 aimed to test whether identity affects norm conformity, independent of any objective reason to do so. We presented participants with opposing norms of an intended ingroup (people the participant is expected to identify with) and outgroup (people the participant is expected not to identify with). If identity does not drive conformity to arbitrary descriptive norms, then these opposite norms should cancel out and we should not see a bias in either direction. On the other hand, self-categorisation theory predicts that participants who more strongly identify with the ingroup over the outgroup will more strongly conform to the arbitrary ingroup norm over the outgroup norm. This experiment was pre-registered at www.aspredicted.org/blind.php?x=hr7cs4.

6.7.1 Method

Participants

180 new English-speaking participants (mean age = 34 years old, 38.3% female), recruited via Amazon's Mechanical Turk, were paid US$0.65 for participating. A bootstrap power analysis based on a pilot sample of 75 participants suggested that a sample of 180 participants would give over 95% power to observe a significant result for the effect of interest (the interaction between which actions the ingroup and outgroup norms favoured, and how much the participant identified with the ingroup over the outgroup).

Procedure

Experiment 5 was the same as Experiment 1 except for two important changes. Firstly, participants were now presented with two norms instead of one. One of these norms was the same ingroup norm that was presented in Experiment 1. The other was a norm of how a supposed outgroup had been allocated to act, where most outgroup members were allocated to the option that fewer ingroup members were allocated to. For example, if a female participant rated high in extraversion and agreeableness and was allocated to an ingroup norm favouring reporting the robber, then she would be shown the following norms:

> In the previous study:

- approximately 75% of female participants that rated high in extraversion and agreeableness were allocated to imagine calling the police and reporting the robber and then rate how they would feel, due to an error in their random allocation.
- approximately 75% of male participants that rated low in extraversion and agreeableness were allocated to imagine doing nothing and leaving the robber alone and then rate how they would feel, due to an error in their random allocation."

To avoid any potential confound due to participants focusing more on the first or second piece of information, we also varied whether the ingroup norm was presented before or after the outgroup norm.

Secondly, we included two questions at the end of the experiment to measure the extent to which participants actually identified with the alleged ingroup (e.g. females who rated high in extraversion and agreeableness) and outgroup (e.g. males who rated low in extraversion and agreeableness). Specifically, we used the single-item social identification measure developed by Postmes, Haslam, and Jans (2013), which asked participants to rate their agreement with the statements "I identify with [INGROUP DESCRIPTION]" and "I identify with [OUTGROUP DESCRIPTION]", on a 7-point Likert scale ranging from "Fully Disagree" to "Fully Agree".

Design and Analysis

A 2 (ingroup norm favours reporting the robber and outgroup norm favours not reporting the robber [$n = 89$] or vice versa [$n = 91$]) x 2 (ingroup norm presented first [$n = 94$] or second [$n = 86$]) between-subjects design was used. Given we only varied whether the ingroup norm was presented first or second to control for a potentially confounding bias, we collapsed across this when analysing the data. Participants who failed the understanding check were excluded from the analysis.

The data were analysed using a preregistered ordinal logistic regression, which predicted responses to the moral dilemma as a function of the following two variables and their interaction:

1. Whether the ingroup norm favoured reporting the robber and the outgroup norm favoured not reporting the robber or vice versa.
2. The difference between how much the participant identified with the supposed ingroup compared to the outgroup. This was calculated by subtracting the participant's rating of identification with the outgroup from their rating of identification with the ingroup.

6.7.2 Results

Figure 6.5 shows the distribution of responses to the moral dilemma for the 121 participants who answered the understanding check correctly, as a function of whether the ingroup norm favoured reporting the robber and the outgroup norm favoured not reporting, or vice versa.

An ordinal logistic regression did not find a significant main effect of whether the ingroup or outgroup norm favoured reporting the robber ($OR = 0.90$, $p = .811$, $95\%CI[0.38, 2.11]$) or a main effect of the extent to which the ingroup was identified with over the outgroup ($OR = 1.04$, $p = .679$, $95\%CI[0.87, 1.24]$)). Crucially, there was a significant interaction between these two variables, wherein participants who more strongly identified with the ingroup over the outgroup conformed more strongly to the ingroup norm rather than the outgroup norm ($OR = 1.30$, $p = .037$, $95\%CI[1.02, 1.67]$)). This result shows that it is the degree to which participants identify with the ingroup that determines the extent to which they conform to random, arbitrary descriptive norms associated with the ingroup, as predicted by self-categorisation theory.

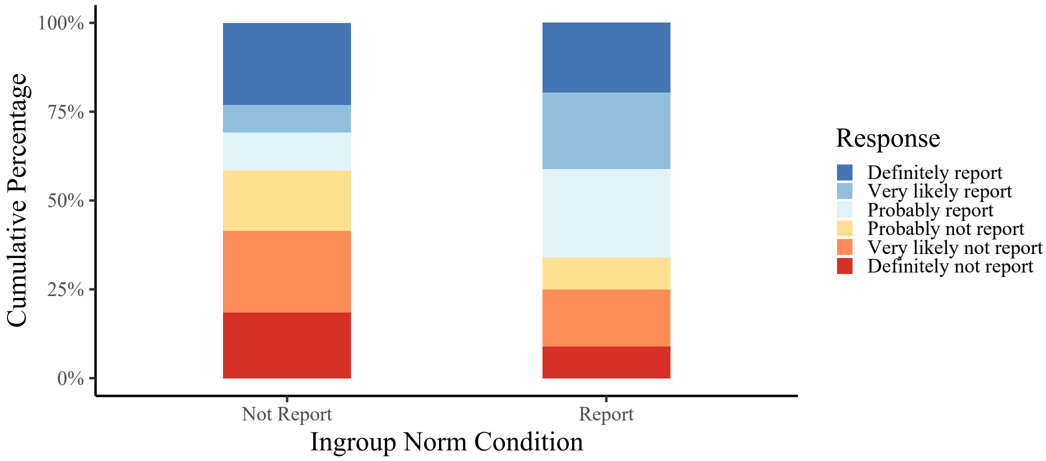

FIGURE 6.5: Proportion of responses in Experiment 5 as a function of whether the ingroup norm favoured reporting the robber (and the outgroup norm favoured not reporting the robber) or vice versa. Responses tended to shift towards the ingroup norm rather than the outgroup norm. Though not apparent from this figure, this effect was strongest for participants who more strongly identified with the ingroup over the outgroup.

6.8 Discussion

This study evaluated whether people conform to random, arbitrary descriptive norms that do not reflect other people's actual preferences. Experiment 1 found that arbitrary descriptive norms influenced participant's preferences during a moral dilemma. Experiment 2 replicated this finding with a different dilemma. Experiment 3 conceptually replicated this finding again and showed that the result continued to hold even when the wording was altered to focus attention on the least popular option, ruling out the possibility that people conformed to the stated norms merely because of increased exposure to that option.

Of course, these results would not be interesting if they arose from our participants being confused and believing that the reported norms actually did convey information about the preferences of previous participants. We addressed this concern in two ways. First, Experiment 4 presented a norm that referred to a different task, so was not relevant to the task at hand. If participants were blindly following whatever norm was presented to them then this norm should have also affected their behaviour. We failed to find any such effect. Second, at the end of each experiment we ran a comprehension check to identify any participants who held this belief. Excluding these participants did not alter the pattern of our results. Nevertheless, there was a notably high failure rate on this comprehension check, suggesting that this task may have been confusing to participants. This is potentially a result of our decision to use a highly controlled but consequently less intuitive task. It would be interesting to see if these results replicate under realistic and incentivised choice scenarios to better understand their real-world significance and reduce any potential for misunderstanding.

Experiment 4 also argues against another potential explanation for our results, that of a basic anchoring process. Anchoring refers to the finding that presenting information at one end of a scale biases judgements towards that information (Tversky & Kahneman, 1974). For example, Tversky and Kahneman (1974) found that people's estimates of various percentages were systematically higher when they had previously spun a high, rather than a low, number on a wheel with numbers 0 to 100. In the case of our arbitrary descriptive norm effect, it could be that participants were anchoring to the norm, thereby shifting their preferences towards the option that was arbitrarily more popular. However, these anchoring processes occur even when the information is clearly irrelevant to the current task (Englich & Mussweiler, 2001; Englich, Mussweiler, & Strack, 2006). An anchoring explanation would therefore expect participants to continue to conform to arbitrary descriptive norms even when we stated that they were irrelevant to the current dilemma in Experiment 4. This is not what we found and thus, the tendency to conform to arbitrary descriptive norms does not appear to be simply due to anchoring.

Finally, Experiment 5 found that participants more strongly conform to arbitrary descriptive norms of groups that they identify with over those they do not identify with. This was demonstrated by the fact that, when presented with two opposite norms, participants who identified with one group over the other group tended to conform to the norm of the group they identified with rather than the group they did not identify with.

This result is consistent with previous findings that also found that the degree to which participants' self-identified with a group determined the degree to which they conformed with the behaviour of that group's members (Neighbors et al., 2010; Rimal, 2008; Smith & Terry, 2003). However, those previous studies presented non-arbitrary norms and, thus, offered objective reasons to more strongly conform when self-identification with the ingroup was high. For example, the informational account could explain a stronger effect of ingroup norms based on the idea that an individual's goals and utilities are more aligned with their ingroup, such that ingroup norms are more relevant to them. Experiment 5 provided evidence that increasing self-identification with the ingroup increases ingroup norm conformity, even when there is no objective reason for it to do so, a finding that the information and social sanction accounts cannot explain. This role of self-identification was successfully predicted by self-categorisation theory, which brings us to the theoretical implications of our findings.

Two of the most prominent explanations of the descriptive norm effect, the informational account and the social sanction account, both assume that people conform to descriptive norms because descriptive norms reflect other people's preferences (Deutsch & Gerard, 1955). Our finding that people conform to arbitrary descriptive norms shows that norms do not have to provide useful information about people's preferences for norm conformity to occur. This result cannot be explained by either the information or social sanction accounts as both accounts predict that people should not conform to arbitrary descriptive norms. This shows that these prominent theories cannot, on their own, provide a complete account of the descriptive norm effect. Our results demonstrate that an additional mechanism is needed and that this additional mechanism is affected by group identification. Our paradigm allows this additional mechanism to be studied, independent of any informational or social sanction effects.

One of the key assumptions of self-categorisation theory is that people tend to internalise the characteristics of salient social groups with which they personally identify (Turner et al., 1987). In typical studies of self-categorisation, the relevant social group is a well-defined group of people that the individual would likely interact with and potentially consider part of their self-identity outside of the confines of the experiment, such as people from your university (e.g. Cruwys et al., 2012; Platow et al., 2005; Terry & Hogg, 1996). In contrast, in our experiments, the relevant social group was not one that a current participant would identify with outside of the experiment. It consisted of people who happened to previously participate in a similar study and were similar to the current participant but who the current participant would likely never interact with. Despite this, we found that participants tended to conform to the descriptive norms of these impersonal, unfamiliar groups. This finding is consistent with the claim of self-categorisation theory that people's identity is not stable but is instead driven by context (Turner et al., 1987). Though these groups may not be how participants typically defined themselves, in the context of our experiments, these groups became salient and so formed the basis for self-categorisation.

In conclusion, this work suggests that people conform to descriptive norms even when they are entirely arbitrary. These results are consistent with the self-categorisation account of norm adherence and cannot be explained by either the informational account or the social sanction account. Our results generalise to different dilemmas and cannot be attributed to participants misunderstanding or misreading the arbitrary descriptive norms. Given the impact that descriptive norms can have on decisions and their increasing application in nudging strategies, the better we understand the mechanisms underlying this normative influence, the better we can design normative nudges to encourage pro-social behaviour.

Chapter 7

Preface to Chapter 8

In the previous chapter we found that people were still influenced by descriptive norms that were completely arbitrary, occurring for random reasons rather than reflecting anybody's actual preferences. This result could not be explained by the informational account or social sanction account of the descriptive norm effect, both of which assume that norms are conformed to because they are indicative of other people's preferences. We offered some initial evidence that these results might be explained by self-categorisation theory.

Self-categorisation theory assumes that people conform to the descriptive norms of salient groups that they identify with (their ingroups) and avoid conforming to the norms of groups they do not identify with (their outgroups, Hogg et al., 1990). Consistent with this, we found that participants presented with two opposing arbitrary descriptive norms, one from an ingroup and one from an outgroup, tended to conform to the ingroup descriptive norm more than the outgroup descriptive norm. However, this was only a weak test of self-categorisation theory. We did not show that people actively avoid conforming to outgroup descriptive norms, only that they are more influenced by ingroup descriptive norms. If we simply assume that ingroup descriptive norms garner more attention (and thus have a stronger effect), this result could be explained without the need for self-categorisation theory.

In response to our paper looking at arbitrary descriptive norms, Reynolds (2019, p. 15) wrote:

> "In their study, Pryor et al. have clarified the theoretical fault lines concerning 'how' norms impact behaviour... The innovative methodology can be extended and perhaps used to examine further the impact of self-categorisation and the ingroup and outgroup dynamics of social influence. This paper represents an important starting point for a trajectory of research examining the underlying mechanisms that account for the impact of norms on behaviour."

Extending this line of research to more directly test a key prediction of self-categorisation theory (that people will avoid conforming to outgroup descriptive norms) is exactly what we do in the next chapter.

Chapter 8

Conformity to the Descriptive Norms of People with Opposing Political or Social Beliefs

The contents of this chapter are adapted from our paper published in *PLOS ONE*: Pryor, C. G., Perfors, A., & Howe, P. D. L. (2019a). Conformity to the descriptive norms of people with opposing political or social beliefs. *PLoS ONE*, *14*(7), e0219464

8.1 Abstract

The descriptive norm effect refers to findings that individuals will tend to prefer behaving certain ways when they know that other people behave similarly. An open question is whether individuals will still conform to other people's choices and behaviour when they do not identify with these other people, such as a Democrat being biased towards following a popular behaviour amongst Republicans. Self-categorisation theory makes the intuitive prediction that people will actively *avoid* conforming to the norms of groups with which they do not identify. We tested this by informing participants that a particular option was more popular amongst people they identified with (an ingroup) and additionally informed some participants that this option was unpopular amongst people they did not identify with (an outgroup). Specifically, we presented descriptive norms of people who supported different political parties or had opposing stances on important social issues. Counter to self-categorisation theory's prediction, we found that informing participants that an option was popular amongst their outgroup led participants' preferences to shift towards that option. These results suggest that a general desire to conform with others may outpower the common ingroup vs outgroup mentality.

8.2 Introduction

Our choices and judgements are influenced by the choices and judgements of other people (Asch, 1951; Sherif, 1936). In particular, we tend to prefer to behave in certain ways when we know that most other people also behave that way (Cialdini et al., 1990). This descriptive norm effect has been successfully used to encourage various prosocial behaviours such as decreasing tax evasion (Wenzel, 2005), decreasing energy use (Schultz et al., 2007) and increasing organ donor registrations (Behavioural Insights Team, 2013), though they can also encourage anti-social behaviour such as corruption (Abbink, Freidin, Gangadharan, & Moro, 2018; Köbis, Van Prooijen, Righetti, & Van Lange, 2015), even in the presence of strong moral pressures against the behaviour (Bicchieri & Xiao, 2009). Given the importance that descriptive norms play in our decisions, it is both theoretically and practically beneficial to understand why and when we conform to descriptive norms.

Self-categorisation theory is a seminal theory of both the descriptive norm effect (Hogg et al., 1990), and a range of other social phenomena such as ingroup favouritism (Tajfel, 1970), stereotyping (Haslam et al., 1996) and power dynamics (Turner, 2005). Self-categorisation theory proposes that an individual's identity is linked to the social groups with which they identify, termed their 'ingroups' (Turner et al., 1987). In an effort to maintain a personal sense of ingroup identity, self-categorisation theory proposes that people will adopt the characteristics of these salient social ingroups, leading to ingroup norm conformity.

Through this emphasis on group identification, self-categorisation theory predicts that the degree to which an individual will conform with the norms of an ingroup will be determined by how strongly they perceive themselves to be a member of that ingroup (Hogg et al., 1990). Consistent with this prediction, Wellen, Hogg, and Terry (1998) found that people were more likely to conform to the norms of a particular group when the salience of their membership with that group was experimentally heightened. This was done by having participants describe how they were similar to other members of the group. Similarly, other studies have found that the more a participant identifies with a group, the more strongly they are influenced by that group's norms (Rimal, 2008; Smith & Terry, 2003; Liu, Thomas, & Higgs, 2019). We even observed this in Experiment 5 of Chapter 6. However, Rimal and Real (2003, 2005) failed to replicate this result, suggesting that group identity may not always determine whether individuals adhere to the ingroup norm.

Furthermore, finding that people conform more with groups that they strongly identify with is not a strong test of self-categorisation theory. Essentially any theory could account for the stronger effect of ingroup norms by assuming that people simply give more attention to normative information from groups they strongly identify with. Giving greater attention to information increases its potential to be incorporated into the decision-making process and thus, affect preferences. The idea that greater attention is given to ingroup norms seems like a reasonable assumption given people tend to attend more to (and better remember) personally relevant information (Celsi & Olson, 1988; Wingenfeld et al., 2006; Breska, Israel, Maoz, Cohen, & Ben-Shakhar, 2011). This underscores the importance of testing other predictions of self-categorisation theory with respect to the descriptive norm effect.

A key prediction of self-categorisation theory is that people will actively avoid conforming to behaviour that is popular or endorsed by groups with which they do not identify (outgroups; Hogg et al., 1990). Just as it predicts that people will conform to salient ingroup norms in order to maintain their sense of ingroup identity, self-categorisation theory predicts that people will avoid conforming to salient outgroup norms for a similar reason; remaining distinct from the outgroup helps people to maintain their sense of ingroup identity. It follows that individuals should conform even more strongly to an ingroup norm when an outgroup tends to behave in the opposite way.

In contrast, explanations of the descriptive norm effect that do not focus on self-categorisation theory typically ignore group identity. Instead, they make the more general assumption that people will follow "what most people do" (Rimal, Lapinski, Cook, & Real, 2005, p. 434). This alternative hypothesis predicts that people will conform to the overall norm, integrating across ingroup and outgroup descriptive norms.

Relevant to the effect of outgroup norms, research into when people begin to diverge from popular behaviour has found that such divergences tend to occur when an outgroup begins adopting the behaviour. Identity-signalling proposes that people will diverge from a behaviour that becomes popular amongst an outgroup in order to prevent people from making undesirable inferences about their identity (Berger & Heath, 2007). Consistent with this, Berger and Heath (2008) found that people were more likely to abandon a behaviour (such as using a catchphrase or wearing a yellow wristband) when the behaviour was adopted by a more dissimilar group.

While this result is consistent with the idea that people will avoid conforming to outgroups, identity signalling is focused on how a decision maker adjusts their behaviours to manage how others perceive them (Berger & Heath, 2007). Thus, the identity signalling literature is focused on outgroup conformity/divergence that is public and expected to have external social consequences (Berger & Rand, 2008; Berger & Heath, 2008; White & Dahl, 2006). In contrast, self-categorisation theory proposes that norms can impact even private behaviours due to a desire to maintain an internal, personal sense of identity (Turner et al., 1987). We are interested in seeing whether people diverge from outgroups even when external social pressures to do so are removed.

In the context of private decisions, we are unaware of any previous studies that have looked at whether ingroup and outgroup descriptive norms have opposite effects on the behaviours people prefer to conform to. However, previous studies have looked at the effects of ingroups vs outgroups for other social phenomena. The concept of ingroups and outgroups was popularised by self-categorisation research into the minimal group paradigm, where it was found that people are biased towards helping an ingroup member over an outgroup member, even when group membership was anonymous and designated based on trivial criteria, such as preference for a particular painting (Tajfel, Billig, Bundy, & Flament, 1971).

More relevant to our current study, Hogg et al. (1990) informed their participants that a small set of four other participants from an outgroup (as defined by having different attitudes to the current participant) supposedly favoured a risky (cautious) option. They found that this led the current participant to predict that their own ingroup would favour more cautious (risky) options. A similar study by Krizan and Baron (2007) failed to replicate this effect. In particular, they did not find that presenting information that a small outgroup was cautious had a significant effect on the decisions of their participants. However, this could be attributed to the fact that the outgroup information was presented briefly amidst 20 minutes of ingroup discussion. Cruwys et al. (2012) found that when participants saw someone from their own university eat a lot of popcorn, they tended to eat more popcorn too and vice versa when they saw the ingroup member eat very little popcorn. This effect slightly reversed when seeing a single outgroup member eat a lot or very little popcorn (as predicted by self-categorisation theory), though this effect was not significant. The extent to which results from these studies would generalise to descriptive norms that describe the typical choices of a larger group is unclear, given small or single-person outgroups provide little information about what behaviour is typical of the broader outgroup.

8.2.1 Aims and hypotheses

The aim of this study was to test the prediction of self-categorisation theory that people's behaviour will shift away from behaviour common amongst an outgroup. This was tested against the alternative hypothesis that people will simply conform to the overall descriptive norm. We tested this by presenting participants with an ingroup descriptive norm favouring a certain option and additionally presenting half of the participants with an outgroup descriptive norm that favoured the alternative option. Self-categorisation theory predicts that, in an effort to remain distinct from the outgroup, participants will conform more to the ingroup descriptive norm when an opposite outgroup descriptive norm is shown. The alternative hypothesis is that people will conform to the overall descriptive norm, such that conformity with the ingroup descriptive norm will decrease when an opposite outgroup descriptive norm is presented.

It is worth noting that in the current paper we do not create and designate participants to ingroups and outgroups but instead draw their attention to pre-existing social categories, specifically based on political partisanship or attitude towards social issues such as feminism, gun control and religion. Groups based on these political and social attitudes are constantly in opposition within the US and play an important role in people's social identity (Greene, 1999).

To pre-empt our results, we found that participants' preferences shifted towards (rather than away from) the behaviour that was popular amongst the outgroup descriptive norm, when it was presented. This occurred even when defining the ingroup and outgroup based on political identity or social issues that participants indicated that they cared about, such as gun control. These results are inconsistent with self-categorisation theory and instead argue for a more general mechanism whereby people tend to conform to both the outgroup and ingroup descriptive norm.

8.3 Experiment 1

8.3.1 Method

Participants

301 participants (M_{age} = 40 years, 60% female) from the United States of America were recruited via Mechanical Turk and paid US$0.65 for participation. The experiment took around 3 minutes to complete and was completed in a web browser. Informed consent was obtained in all experiments reported here. Ethics approval for all experiments reported here was granted by the University of Melbourne Human Research Ethics Committee. All experiments were carried out in accordance with the relevant guidelines.

Procedure

After providing basic demographic information about age and sex, participants were asked to select which out of a set of nine topical social issues, such as gun control and immigration, they cared most about. After selecting the issues they cared about most, participants were presented with a statement about their chosen issue (the full list of these statements is contained in Appendix D on page 157). For example, if a participant selected gun control as the issue they cared about most then they were presented with the following statement "Adults should have the right to carry a concealed handgun". Participants were asked to report the extent to which they agreed or disagreed with the statement on an 11-point Likert scale ranging from -5 (Strongly Disagree) to +5 (Strongly Agree). Their rating of this chosen social issue was then used to define the ingroup and outgroup when subsequently presenting descriptive norms, as outlined below. For an example transcript of the information presented to participants, see Appendix E on page 159.

Participants were then presented with instructions for the current study. They were told that this study was following on from a previous study that investigated how people feel during a moral dilemma. This background story was included simply to justify the source of the descriptive norms that were later presented. Participants were told that they would be presented with a scenario describing a moral dilemma and have to choose which action they would take and then rate how they would feel about it.

After reading these instructions, participants were presented with the following moral dilemma:

> Imagine you have witnessed a man rob a bank. However, you then saw him do something unexpected with the money. He donated it all to a run-down orphanage that would benefit greatly from the money. You must decide whether to call the police and report the robber or do nothing and leave the robber alone.

Below this moral dilemma, participants were presented with an ingroup descriptive norm informing them that 60% of previous participants who had agreed with them about their chosen social issue (i.e. members of their political ingroup) chose to act a certain way. Half of the participants were told that their ingroup members mostly chose to "call the police and report the robber" while the remaining half were told that their ingroup members mostly chose to "do nothing and leave the robber alone". So, for example, if participant X indicated that they cared most about gun control, they might have been told that "approximately 60% of participants who agreed with you about gun restrictions chose to call the police and report the robber".

Additionally, half of the participants were also informed that, in the previous study, 85% of participants that disagreed with them on that issue chose the other option. From the example above, participant X would have been informed that "approximately 85% of participants who disagreed with you about gun restriction chose to do nothing and leave the robber alone". A stronger outgroup norm (85%) relative to the ingroup norm (60%) was used because we were interested in the effect of the outgroup norm rather than the ingroup norm. Presenting the stronger outgroup norm was intended to allow participants to be affected by the outgroup norm without the ingroup norm dominating decisions.

Thus, our study had a 2 x 2 balanced design: half our participants were told that the ingroup norm favoured one action whereas the remaining participants were told that it favoured the other action; half our participants were presented only with the ingroup norm, whereas the remaining participants were presented with both the ingroup and outgroup norms. For an example transcript, please see Appendix E on page 159.

Participants then indicated how they would respond to the moral dilemma on a 6-point Likert scale ranging from "Definitely call the police and report the robber" to "Definitely do nothing and leave the robber alone". To fit with the backstory presented in the instructions, participants were also asked to rate how good or bad they felt about their chosen action, although these responses were not analysed.

In order to ensure that participants were paying attention, as is especially recommended for Mechanical Turk studies (Kittur et al., 2008), we included an understanding check asking participants which of the following options was true about the previous study described in the instructions:

1. Participants chose which action they preferred (correct)
2. Due to a computer error, participants were not allocated equally to imagine performing the different actions (incorrect)
3. No data was saved during the experiment. (incorrect)
4. The participants completed the experiment with their eyes closed (incorrect)

Finally, Postmes et al.'s (2013) single-item social identification measure was included after the understanding check to test whether individuals identified with the relevant ingroup and did not identify with the relevant outgroup. This measure simply involves asking participants the extent to which they agree with two statements about whether they identified with the designated ingroup and outgroup. The statements were "I identify with [INGROUP]" and "I identify with [OUTGROUP]", where [INGROUP] and [OUTGROUP] were replaced with the appropriate descriptions (e.g. "Pro-Gun Enthusiasts" and "Anti-Gun Advocates").

Design

A 2 (INGROUP DESCRIPTIVE NORM: 1= reporting vs vs -1 = not reporting) x 2 (NORMS SHOWN: 0 = ingroup only vs 1 = opposing norms) between-subjects design was used. The independent variable NORMS SHOWN refers to whether only an ingroup descriptive norm was shown (ingroup only) or both an ingroup descriptive norm and an outgroup descriptive norm were shown (opposing norms). The variable INGROUP DESCRIPTIVE NORM refers to whether the ingroup descriptive norm favoured reporting the robber (reporting condition) or leaving the robber alone (not reporting condition). When both the ingroup and outgroup descriptive norms were shown, we randomly varied their ordering. This was done only to control for potential order effects and so was ignored when analysing the data. The dependent variable was participants' responses on the Likert scale rating the certainty with which they would act a certain way. Any participants that failed the understanding check were excluded because this indicated that they had not paid attention throughout the task. Additionally, any participants that reported being neutral about their chosen social issue was excluded because this prevented us from determining an ingroup and outgroup.

8.3.2 Results

37 participants were excluded from the analysis for either failing the single-question understanding check (n = 23) or rating their attitude towards their chosen social issue as neutral (n = 14). The distribution of responses for the remaining 264 participants is shown in Figure 8.1 on the next page.

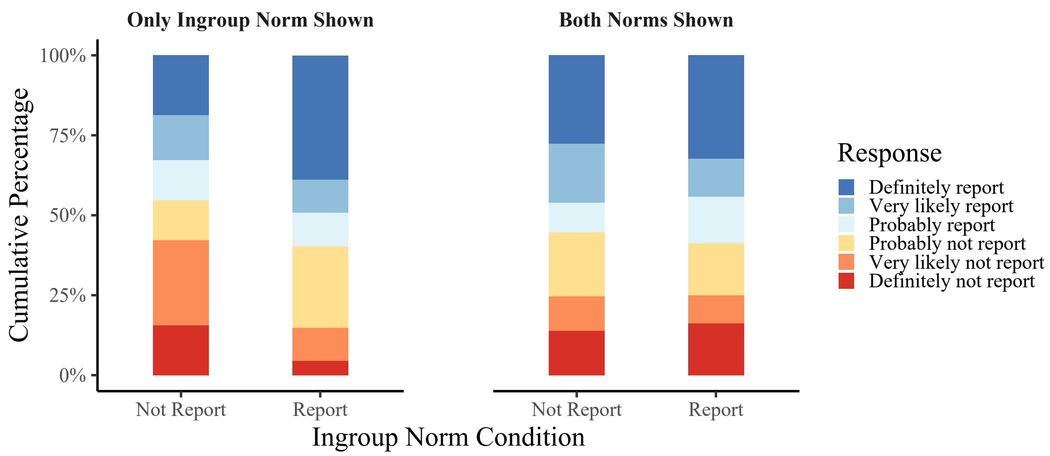

FIGURE 8.1: Distribution of responses by condition in Experiment 1. Responses conformed more with the ingroup descriptive norm when only the ingroup descriptive norm was presented than when an opposing outgroup descriptive norm was also presented, counter to self-categorisation theory's prediction.

Model comparison

In order to ascertain the extent to which self-categorisation theory provides a better or worse explanation of our data than the alternative hypothesis, we directly compared these competing accounts using models that reflect their different assumptions, as described below.

Each of the following models represent Bayesian versions of ordinal logistic regression, which predicts the proportions of responses on an ordinal scale while assuming that certain variables (in our case, the descriptive norms) change the odds of making higher or lower responses on the scale. Specifically, the variables are parameterised in terms of the natural log odds of favouring a higher response (more strongly preferring to not report the robber). We can represent this as shown in Equation (8.1):

$$log_e(\text{odds of responding higher}) = b_{in}I + b_{both}B + b_{out}I \times B \tag{8.1}$$

Here I represents the INGROUP DESCRIPTIVE NORM condition and B represents the NORMS SHOWN condition. Because the option that was more popular with the ingroup was always less popular with the outgroup, we can represent both the presence and direction of the OUTGROUP DESCRIPTIVE NORM based on the interaction between the two independent variables ($I \times B$). Meanwhile, b_{in}, b_{both} and b_{out} are parameters representing the effects of changing these conditions. The self-categorisation and the alternative account make different assumptions about these parameters, which we represent using different priors, as outlined shortly.

One additional piece of data that the self-categorisation model can make use of is whether the participants reported identifying with the ingroup and reported not identifying with the outgroup. According to self-categorisation theory, individuals should only want to conform to a supposed ingroup norm if they actually self-identify with that group and should only want to avoid following an outgroup norm when they consider that outgroup as separate from their self-identity. We thus include these interactions in the self-categorisation model, as outlined in Equation (8.2), where INGROUP AGREE is coded as 1 if the participant reports identifying with the ingroup and 0 otherwise. Equivalent binary coding is used for OUTGROUP DISAGREE.

$$log_e(\text{odds}) = b_{in} I \times \text{INGROUP AGREE} + b_{both} B + b_{out} I \times B \times \text{OUTGROUP DISAGREE} \tag{8.2}$$

The INGROUP AGREE and OUTGROUP DISAGREE variables effectively act as switches, determining whether the self-categorisation model assumes the participant will be affected by the ingroup and outgroup descriptive norms respectively. The alternative hypothesis assumes that identification with the group that a descriptive norm comes from does not influence the effect of that descriptive norm and thus, these variables are ignored by the alternative model. Out of the 264 participants included in this analysis, 219 (83%) identified with the ingroup and 220 (83%) did not identify with the outgroup. Scores on these two variables had no bearing on a participant's inclusion in the data analysis, ensuring that the same participants are analysed for both the self-categorisation model and the alternative model.

Prior assumptions: $\mathbf{b_{in}}$ Given the similarity of our ingroup norm condition to the experiments reported in Chapter 6, we used the observed effect of the ingroup descriptive norm in those experiments to inform our prior for the ingroup descriptive norm effect in the current analysis. The log odds ratio estimated across the relevant experiments from Chapter 6 was 1.02 with a standard deviation of 0.19 (see Appendix F on page 161). A notable difference between those experiments and the descriptive norms that were presented in the current experiment is that the previous experiments presented a stronger ingroup descriptive norm (75% of ingroup did X) than the current experiment (60% of ingroup did X). To adjust for this difference and account for increased uncertainty in this parameter estimate, we set the prior distribution for the effect of the ingroup descriptive norm to be a folded normal distribution with a mean of $\frac{0.6}{0.75} \times 1.02 = 0.816$, and a standard deviation of 0.5 for both the self-categorisation and alternative models. We folded this normal distribution such that the prior is restricted to be greater than 0, given both models assume that people's preferences should shift towards the ingroup descriptive norm (i.e. the effect of the ingroup descriptive norm will be positive).

Prior assumptions: $\mathbf{b_{both}}$ The parameter b_{both} represents a possible main effect on responses of merely presenting both an ingroup and outgroup descriptive norm compared to only an ingroup norm, independent of the direction of those norms. Including this effect in the models is important as it allows for the possibility that the outgroup descriptive norm is more effective in one direction than in another. It also allows for the possibility that being presented with two opposing descriptive norms shifts people's bias. For example, the increased ambiguity caused by having opposing norms presented may elicit an omission bias (Ritov & Baron, 1990), wherein taking no action (i.e. not reporting the robber) is favoured more often, independent of the actual direction of the descriptive norms. Given that the presentation of the outgroup descriptive norm was independent of the direction of either norm, we had no clear, theoretical reason to predict a strong systematic effect in either direction due to merely presenting an outgroup descriptive norm, independent of that descriptive norm's direction. Thus, the self-categorisation and the alternative model both adopt a weakly informative prior for b_{both}, represented by a normal distribution with a mean of 0 and standard deviation of 0.5.

Prior assumptions: $\mathbf{b_{out}}$ The parameter b_{out} is the key manner in which the self-categorisation explanation of the descriptive norm effect differs from that of the alternative hypothesis that people conform to the overall descriptive norm. This parameter represents the extent to which presenting an outgroup descriptive norm that is opposite to the ingroup descriptive norm shifts preferences towards or away from the option favoured by the ingroup descriptive norm.

For self-categorisation theory, presenting an outgroup descriptive norm that opposes the ingroup descriptive norm is expected to increase conformity with the ingroup norm. We represented this with a normally-distributed prior that was restricted to be greater than 0 (specifically a half-normal distribution with a mean of 0 and standard deviation of 0.5).

Contrasting with self-categorisation theory, the alternative hypothesis assumes that people care about the overall descriptive norm, regardless of whether it comes from an ingroup or outgroup. Thus, the ingroup and outgroup descriptive norms are assumed to affect preferences equivalently, only differing to the extent that the strength of these norms differ. Given that 85% of the outgroup was said to have favoured a particular option, compared to 60% of the ingroup favouring the alternative option, we represented this expectation by setting b_{out} to be a transformation of b_{in}, such that $b_{out} = -0.85/0.6 b_{in}$.

Bayes factor We assessed the relative evidence for the self-categorisation model and the alternative model provided by the data using a Bayes Factor (BF) calculated with the "Bridge Sampling" package in R (Gronau et al., 2017). This Bayes Factor represents the probability of the observed data occurring under the alternative model relative to the probability of the observed data occurring under the self-categorisation model. Specifically,

$$BF = \frac{p(\text{data} \mid \text{alternative})}{p(\text{data} \mid \text{self-categorisation})} \tag{8.3}$$

We found a BF of 34.97, suggesting that the observed data was 34.97 times more likely under the alternative model than under the self-categorisation model. This represents strong evidence in favour of the alternative model over the self-categorisation model. These results remain qualitatively the same across different values for the mean of the b_{out} prior (0, 0.5 and 1) and across differing values of the standard deviation of this prior (0.25, 0.5, 1 and 2), with the Bayes Factor always favouring the alternative model and ranging from 30.04 to 522.62.

Figure 8.2 shows the prior and posterior distribution of the parameters for each model. Even when we presented descriptive norms from ingroups and outgroups defined on issues that participants cared about and accounted for whether participants actually identified as ingroup members and did not identify with the outgroup (i.e. under conditions strongly favouring the self-categorisation model), we did not find data consistent with self-categorisation theory. Instead, these results support the alternative hypothesis that people tend to favour the overall descriptive norms, whether they identify with the people this norm relates to or not.

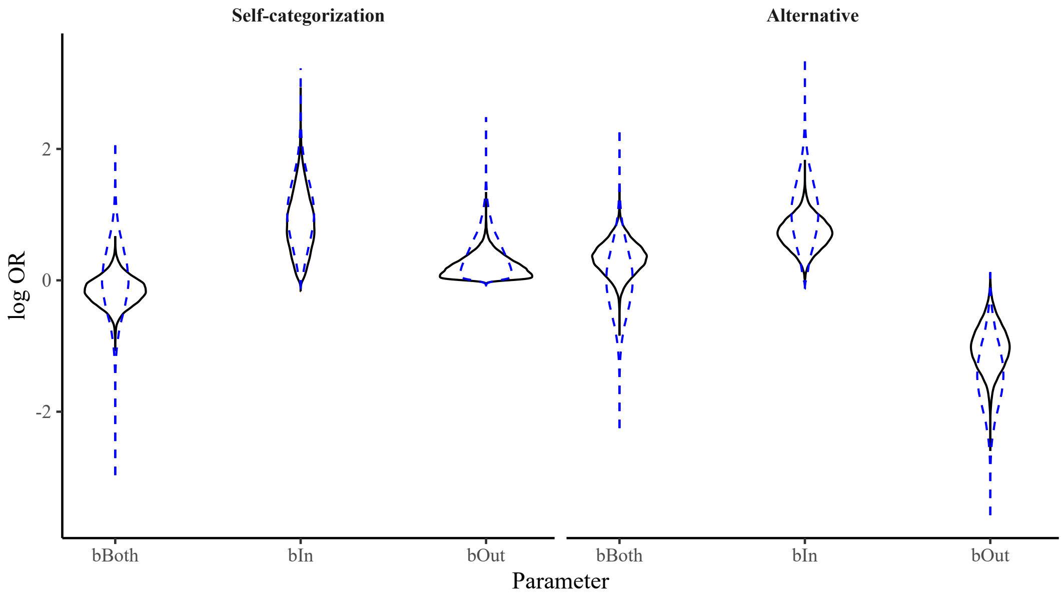

FIGURE 8.2: Violin plots of the prior (dashed blue lines) and posterior (solid black lines) density for each of the parameters in both the self-categorisation model and the alternative model for Experiment 1. The results favour the alternative model over the self-categorisation model.

Effect sizes

As an additional analysis, we ran a frequentist ordinal logistic regression to measure the effect sizes for the parameters reported in Equation (8.1) on page 92. We found a significant effect of the INGROUP DESCRIPTIVE NORM (N = 264, OR = 2.48, 95%CI[1.35, 4.58]), wherein participants' preference shifted towards favouring the option that was popular under the ingroup descriptive norm. There was no significant shift in bias for participants who had opposing norms shown (n = 264, OR = 1.39, 95%CI[0.760, 2.55]). The observed interaction between INGROUP DESCRIPTIVE NORM and NORMS SHOWN (i.e. the effect of the OUTGROUP DESCRIPTIVE NORM) was in the direction consistent with the alternative hypothesis however, this effect was not significant (N = 264, OR = 0.429, 95%CI[0.181, 1.01]).

8.4 Experiment 2

Experiment 1 found evidence against self-categorisation theory, instead favouring the alternative hypothesis that people care more about the overall descriptive norm than about actively not conforming to outgroup descriptive norms. However, a frequentist analysis failed to reject the null hypothesis for the effect of the outgroup descriptive norm. Additionally, it may be the case that self-categorisation mechanisms were not engaged due to the abstract nature of how we defined ingroups and outgroups. Self-categorisation mechanisms may require a clearer social entity with which people can identify, such as political groups. We therefore conducted a pre-registered conceptual replication of Experiment 1, using political identity to define ingroups and outgroups, in order to ascertain whether the results from Experiment 1 were reliable. The pre-registration for this experiment is available at: www.osf.io/j9ugv

8.4.1 Method

Participants

600 English-speaking participants (M_{age} = 40 years, 47% female) from the United States of America agreed to participate via Mechanical Turk in exchange for US$0.65.

8.4.2 Procedure and design

Experiment 2 was exactly the same as Experiment 1 except that participants no longer selected and rated social issues that they cared about. Instead, participants were asked to rate their attitudes towards the two major US political parties (Republican and Democratic party) on an 11-point Likert scale ranging from "Strongly dislike" (-5) to "Strongly like" (+5). These ratings were used to define the ingroup and outgroup when subsequently presenting descriptive norms. Specifically, if the participant reported liking the Democratic party and disliking the Republican party, then the ingroup was Democratic party supporters while the outgroup was Republican party supporters and vice versa for those that reported liking the Republican party. For the 11 participants who reported disliking both the Republican and Democratic party, the ingroup was designated as Independents or other party supporters and their outgroup was designated as Republican or Democratic party supporters.

This was reversed for the five participants that reported liking both parties. Twenty-five participants were excluded for rating both parties as neutral, preventing allocation of ingroups and outgroups.

8.4.3 Results

92 participants were excluded from the analysis for either failing the understanding check (n = 53) and/or rating both political parties as neutral (n = 42), preventing determination of an ingroup and outgroup. The distribution of responses for the remaining 508 participants is shown in Figure 8.3. Out of these participants, 428 (84%) identified with the ingroup and 437 (86%) did not identify with the outgroup, suggesting the designation of ingroups and outgroups was generally, though not universally, successful.

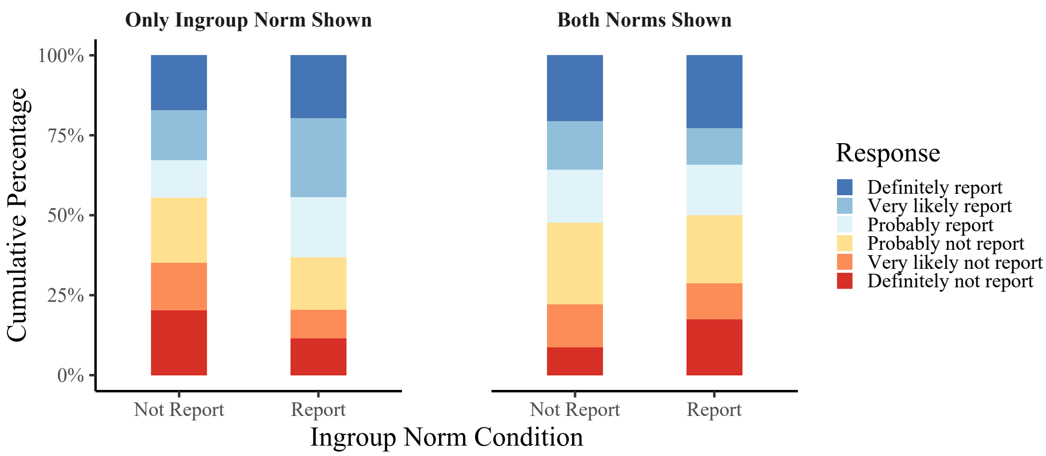

FIGURE 8.3: Distribution of responses by condition in Experiment 2. Responses conformed more with the ingroup descriptive norm when only the ingroup descriptive norm was presented than when an opposing outgroup descriptive norm was also presented, counter to self-categorisation theory's prediction.

An ordinal logistic regression found a significant effect of the INGROUP DESCRIPTIVE NORM (N = 508, OR = 1.79, 95%CI[1.16, 2.78]), wherein participants' preferences tended to favour the ingroup descriptive norm. There was no significant shift in bias for participants who had opposing norms shown (N = 508, OR = 1.41, 95%CI[0.91, 2.18]). The observed interaction between INGROUP DESCRIPTIVE NORM and NORMS SHOWN (i.e. the effect of the OUTGROUP DESCRIPTIVE NORM), found that preferences significantly shifted towards the outgroup descriptive norm and away from the ingroup descriptive norm, consistent with the alternative hypothesis (N = 508, OR = 0.477, 95%CI[0.26, 0.88]).

Direct comparison of the self-categorisation model and the alternative model found a BF of 286.31 in favour of the alternative model, showing strong evidence against the self-categorisation model. Figure 8.4 on the following page shows the prior and posterior distribution of the parameters for each model. Even when we presented descriptive norms allegedly associated with political identity, participants' responses tended to shift towards, rather than away from, the descriptive norm of the outgroup.

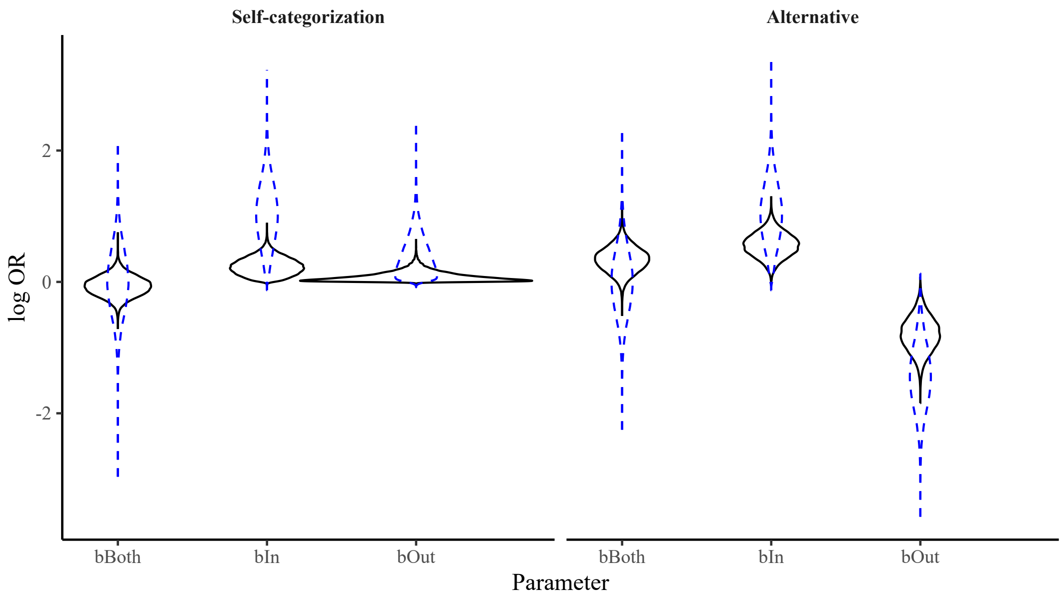

FIGURE 8.4: Violin plots of the prior (dashed blue lines) and posterior (solid black lines) density for each of the parameters in both the self-categorisation model and the alternative model for Experiment 2. The results favour the alternative model over the self-categorisation model.

8.5 Experiment 3

The previous two experiments presented evidence against self-categorisation theory, finding that people's preferences shift towards, rather than away from, an outgroup descriptive norm. In Chapter 6, we provided evidence against the two other most prominent explanations of the descriptive norm effect, the informational account and the social sanction account. These two theories assume that descriptive norms provide useful information about the utility of the options (Cialdini et al., 1990; Deutsch & Gerard, 1955) or the likely social consequences of violating a norm (Ajzen & Fishbein, 1980; Bendor & Swistak, 2001; Deutsch & Gerard, 1955). As a result, they cannot account for our finding in Chapter 6 that people follow arbitrary norms that offer no useful information.

Taken together, the results from this and the previous chapter provide evidence that none of self-categorisation theory, the informational account or the social sanction account can fully explain descriptive norm effects. However, it remains possible that self-categorisation was driving the effects of arbitrary descriptive norms observed in Chapter 6 while informational or social sanction mechanisms were driving the effects of the ingroup and outgroup descriptive norms presented in this chapter. For example, following the informational account, we may assume that ingroup and outgroup members are both sources of useful information as to which option is best.

Therefore, this account could explain why we found that people shifted towards an outgroup descriptive norm when it was presented.

To simultaneously test these three accounts, Experiment 3 presents ingroup and outgroup descriptive norms that are entirely arbitrary and do not reflect other people's preferences (see Chapter 6). Conformity to these arbitrary norms cannot be explained by the informational or social sanction account. Meanwhile, if participants' preferences continue to shift towards, rather than away from, the outgroup descriptive norm, this would be inconsistent with self-categorisation theory. Consequently, if we replicate the results from the previous two experiments but with arbitrary descriptive norms, this would indicate that none of the three most prominent accounts of the descriptive norm effect can fully account for it. As a result, new explanations would be needed.

8.5.1 Method

Participants

308 English-speaking participants (M_{age} = 40 years, 51% female) from the United States of America were recruited via Mechanical Turk and were paid US$0.65 for participation.

Procedure and design

Experiment 3 was exactly the same as Experiment 1 except that now instead of presenting a descriptive norm that allegedly reflected people's preferences, we instead presented a descriptive norm that allegedly occurred due to an error-prone, random-allocation process.

Specifically, the instructions informed participants that we were following up on a previous study that looked at how people feel during a moral dilemma. However, they were told that this previous study did not allow individuals to choose how they would act during the moral dilemma but instead randomly allocated them to imagine performing a particular action and then asked how good or bad they would feel about performing this action. We informed our participants that there was a computer error in the previous study that meant that participants in that study were not allocated equally to the different actions.

The descriptive norms were updated accordingly. For example, an ingroup descriptive norm might have now been:

"In the previous study, approximately 60% of participants who agreed with you about gun restrictions were randomly allocated to imagine calling the police and reporting the robber, due to an error in their random allocation"

The understanding check was also updated as follows:

We were following up on a previous study in this task. Given what we described in the instructions, which of the following was a problem with the previous study?

1. Due to a small computer error, participants were not allocated equally to imagine performing the different actions. (correct)
2. One action was preferred by most participants, while very few participants said they were likely to perform the other action. (incorrect)
3. No data was saved during the experiment. (incorrect)
4. The participants completed the experiment with their eyes closed. (incorrect)

8.5.2 Results

185 participants were retained in the analysis after excluding participants that failed the understanding check and participants that did not indicate an opinion on the statement about their chosen social issue (i.e. they rated the statement as neutral), which prevented allocation of an ingroup and outgroup. The distribution of responses from these participants for each condition is shown in Figure 8.5.

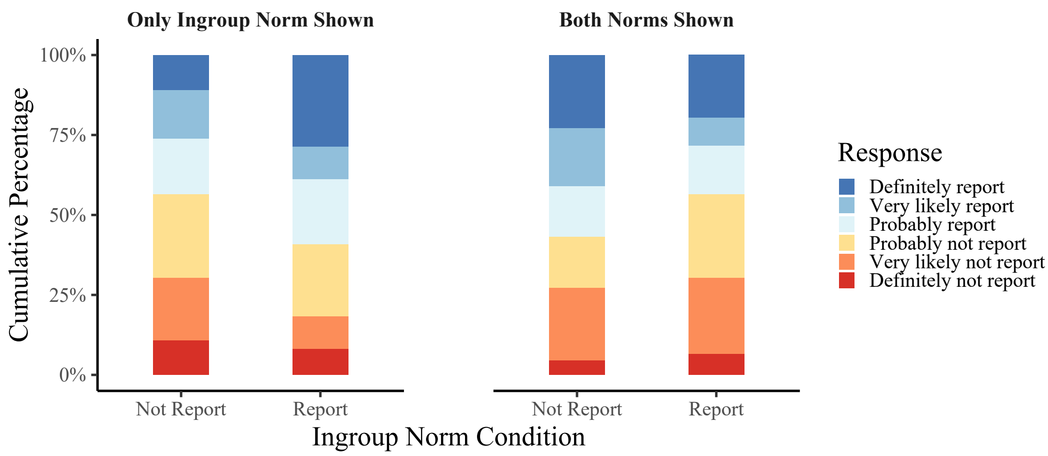

FIGURE 8.5: Distribution of responses by condition in Experiment 3. Responses conformed more with the arbitrary ingroup descriptive norm when only the ingroup descriptive norm was presented than when an opposing outgroup descriptive norm was also presented. This result is inconsistent with self-categorisation theory, the informational account and the social sanction account.

An ordinal logistic regression found that participants tended to follow the INGROUP NORM but this effect did not reach significance ($N = 185$, $OR = 1.97$, $95\%CI[0.969, 4.03]$). There was not a significant main effect of having opposing norms compared the only the ingroup norm ($N = 185, OR = 1.72, 95\%CI[0.841, 3.54]$). Of key importance, and consistent with the alternative hypothesis, the interaction between INGROUP NORM and NORMS SHOWN (i.e. the effect of the OUTGROUP NORM) was significantly lower than 1 ($N = 185$, $OR = 0.346$, $95\%CI[0.123, 0.964]$). Thus, presenting an outgroup norm that was opposite to the ingroup norm shifted preferences away from the ingroup norm. Self-categorisation theory cannot explain this result.

Direct comparison of the self-categorisation model and the alternative model found a BF of 80.48 in favour of the alternative model, showing strong evidence against the self-categorisation model. Figure 8.6 shows the prior and posterior distribution of the parameters for each model. Even when we presented arbitrary descriptive norms, participants' responses tended to shift towards, rather than away from, the descriptive norm of the outgroup. If we collapse across all of the experiments from this study, we find a Bayes Factor of 13,917 in favour of the alternative model.

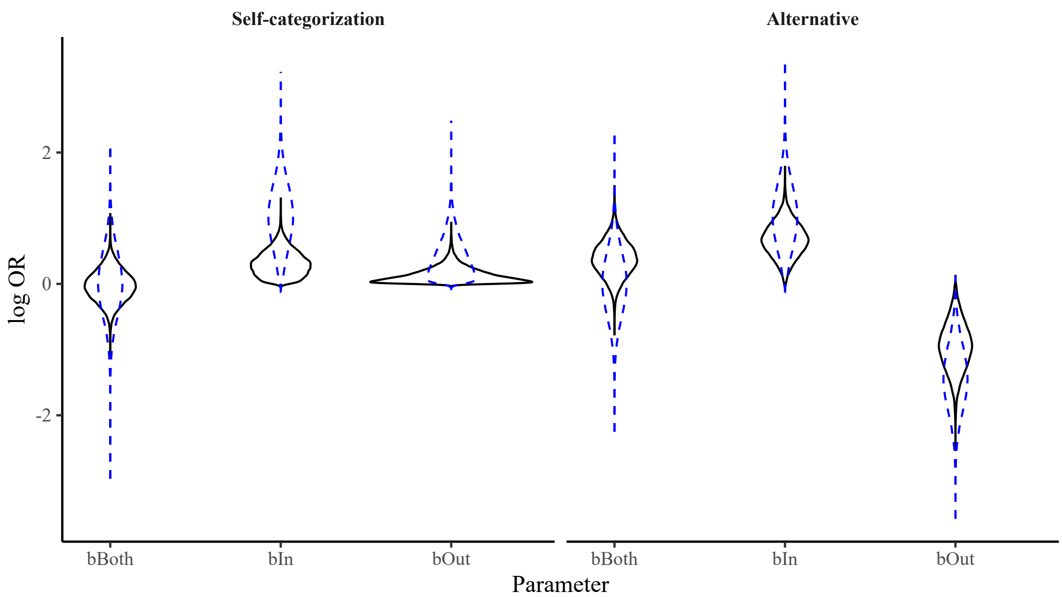

FIGURE 8.6: Violin plots of the prior (dashed blue lines) and posterior (solid black lines) density for each of the parameters in both the self-categorisation model and the alternative model for Experiment 3. The results favour the alternative model over the self-categorisation model.

8.6 Discussion

According to self-categorisation theory, people are expected to actively conform to the norms of groups with which they identify at a given time (ingroups) and actively try to avoid conforming to the norms of groups with which they do not identify (outgroups, Hogg et al., 1990). Across three experiments, we found evidence against this; participants shifted their preferences away from an option that was popular with an ingroup when they were additionally presented with an outgroup descriptive norm that favoured the alternative option. This was found even though ingroups and outgroups were defined based on salient, important issues such as political identity or social issues that participants themselves indicated that they personally cared strongly about and identified with (e.g. gun control). Thus, our findings provide strong evidence that self-categorisation theory is unable to fully explain the descriptive norm effect. Instead, it may be that an option being popular, even amongst an outgroup, can increase preferences towards that option.

In addition to arguing against self-categorisation, we also found that our results could not be explained by the informational or social sanction accounts of the descriptive norm effect. Chapter 6 already found evidence against these more objective mechanisms by showing that people are influenced by arbitrary descriptive norms that offer no useful information. However, it could have been that self-categorisation was influencing the results in Chapter 6 while informational or social sanction mechanisms were driving the impact of the non-arbitrary norms presented in Experiments 1 and 2. Therefore, Experiment 3 continued to look at the effect of outgroup descriptive norms but using the arbitrary descriptive norms from Chapter 6. We found that even with arbitrary ingroup and outgroup descriptive norms, participants showed the same pattern of results. The fact that our participants were influenced by arbitrary descriptive norms means that the informational and social sanction accounts cannot explain our results. Meanwhile, participants' preferences continued to shift towards, rather than away from, the outgroup descriptive norm, arguing against self-categorisation theory. Thus, none of the three most prominent accounts of the descriptive norm effect can account for the results reported in this chapter.

It is worth emphasising that our results do not mean that self-categorisation never influences the descriptive norm effect. The fact that we found results counter to self-categorisation theory, even under conditions theoretically suited to it, suggests that self-categorisation is less general than previously thought and cannot fully account for the descriptive norm effect. However, it may be that self-categorisation contributes to the descriptive norm effect but only under more restricted conditions. For example, in the current paper, the descriptive norms were experimentally varied such that they had no pre-existing association with either the ingroup or outgroup. It may be that self-categorisation only applies when a descriptive norm is strongly stereotypical of a group, such that it is a well-established part of ingroup or outgroup identity. Testing self-categorisation theory under such conditions would be impractical though, given experimentally manipulating people's ongoing stereotypes is difficult and likely unethical.

Alternatively, self-categorisation mechanisms may be limited to when the descriptive norms are stronger. It may be that if a behaviour or preference is not essentially uniform across a group (i.e. approximately 100% of the group engages in it) then it is not considered a salient part of that group's identity. In the current paper, we used descriptive norms that were weaker than 100%. Perhaps presenting norms that are closer to uniform might be considered more relevant to group identity and thus, self-categorisation mechanisms might be more engaged. Given most descriptive norms, both appearing in research and the real-world, are generally weaker than 100%, this would substantially limit the relevance of self-categorisation theory. Furthermore, there is some evidence that stronger descriptive norms have a weaker effect on people's preference (Czajkowski et al., 2019, e101110), which runs counter to this possibility. Nevertheless, further studying the functional association between the strength of descriptive norms and normative influence would be a compelling area for future research.

One other factor that may have restricted any potential desire to avoid conforming to the outgroup comes from the identity signalling literature. Identity signalling relates to the idea that people diverge from behaviour that is popular amongst an outgroup to avoid people mistakenly identifying them as members of that group (Berger & Heath, 2007). Crucially, this research suggests that people will only diverge from behaviours that are relevant to the individual's identity (Berger & Heath, 2008). While identity signalling is focused on social consequences to public behaviours (which do not apply to the private decisions studied here), this notion of identity relevance may still apply.

Identity signalling research has shown that certain products (e.g. clothing or cars) are considered more relevant to identity than other products (e.g. soap or bike lights). While it seems reasonable to believe that moral decisions should be relevant to group identity defined by ideological stances (i.e. political association and attitude towards social dilemmas), other features may be required. Existing research suggests that behaviours tend to be particularly relevant to identity when there are a large number of available options that people can distinguish themselves on (Belk, 1981), when the choice takes effort to make (Belk, 1981), when the behaviour is publicly visible (Berger & Heath, 2007; Berger & Rand, 2008; White & Dahl, 2006) and when the decision does not revolve around function (Shavitt, 1990). The moral dilemmas used in the current study do not satisfy these features well. In particular, there were only two options (limiting the scope for group differentiation) and the decisions were private. It would therefore be interesting to see whether our findings extend to decisions involving options that have been verified as highly identity relevant, such as clothing.

Alongside not claiming that self-categorisation theory never applies, we are also not arguing that people do not distinguish between the ingroup and the outgroup. The alternative model specified in this paper made a simplifying assumption that ingroup and outgroup descriptive norms would have equivalent effects on people's preference (assuming they are of equal strength). This is an unrealistic assumption and likely served to hinder the alternative model. Experiment 5 in Chapter 6, for example, presented norms from an ingroup and outgroup of equal strength and found that participants tended to conform more to the ingroup norm than the outgroup norm. While people's responses to ingroup and outgroup norms may tend to differ in degree, our results suggest that they do not differ in direction.

In terms of the generality of this finding, it is worth noting that we focused exclusively on descriptive norms, relating to what most other people do. Another important social pressure involves how most other people feel we ought to behave, termed injunctive norms (Cialdini et al., 1990). Descriptive and injunctive norms do not always align (Park, Lee, & Han, 2007) and may have distinct effects (Melnyk, Herpen, Fischer, & van Trijp, 2011). Given injunctive norms more directly relate to the attitudes of other people, it is possible that they are more relevant to group identity and thus, may be more likely to engage self-categorisation mechanisms. Nevertheless, it has been found that people tend to automatically and implicitly infer the existence of injunctive norms when informed about descriptive norms (Eriksson et al., 2015), so it is possible our participants were not only influenced by our explicitly presented descriptive norms but also by their inferred injunctive norms.

One potential factor contributing to our observed shift away from the ingroup norm when an outgroup norm was presented is an increase in cognitive deliberation. Receiving conflicting normative information makes it harder to rely on these norms in order to make a decision. For example, when only an ingroup norm is presented, participants could heuristically favour the option that was indicated as being more popular. Thus, it could be that when the outgroup norm conflicted with the ingroup norm, participants were forced to think more carefully about which option they preferred. If this increase in cognitive deliberation drove the effect of the outgroup norm, we would expect conformity towards either norm to regress towards 50% when the outgroup norm was present. The fact that preferences shifted past 50% (with more participants conforming to the stronger outgroup norm) suggests that a decreased reliance on norms due to cognitive deliberation alone cannot explain our findings. Instead, it seems that some participants were conforming towards the option strongly favoured by the outgroup, despite reporting that they did not identify with that group.

Our data shows that, at least in some circumstances, people tend to conform to descriptive norms even when they are exhibited by an outgroup. There are a number of reasons why this might occur. One possible reason is that people may have an evolved or conditioned anxiety response to deviating from the overall descriptive norm. If everyone else is engaging in a behaviour, it is often because that behaviour has been deemed beneficial by others (Cialdini et al., 1990) and deviating from it may cause you to stand out and expose yourself to additional risk (Hamilton, 1971). Psychological studies (Darley, 1966; Sarnoff & Zimbardo, 1961), sociological studies (Gelfand et al., 2011) and agent-based simulations (Roos, Gelfand, Nau, & Lun, 2015) have shown that conformity or societal norms tend to be stronger when the society or individuals within it are threatened. Thus, people may conform to norms simply because they feel anxious otherwise.

Alternatively, it may be that some more general mechanisms that are typically applied when explaining other choice biases could also explain the descriptive norm effect. We discussed a number of these more general mechanisms in Chapter 2.

Regret aversion is one theory that is commonly applied to other choice biases, such as the omission bias.

Studies have found that people experience more regret for bad outcomes that result from actions rather than simply from omitting to act (Baron & Ritov, 1994; Kahneman & Miller, 1986; Nicolle et al., 2011). Regret aversion thus proposes that people are biased to omissions over acts in order to reduce the potential negative experience of regret. In the case of descriptive norms, it may be that people can offload the sense of responsibility for a decision if most other people are engaging in that behaviour too. Acting differently to how most people are behaving requires taking on a greater level of personal responsibility for the decision, increasing the potential for anticipated regret.

Consistent with the idea that anticipated regret is reduced when offloading the responsibility to others, Steffel and Williams (2018) found that regret decreased when delegating a decision to someone else. This occured even when the outcome of the decision was the same and that delegate was inexperienced. Similarly, this diffusion of responsibility may also decrease the level of guilt people feel when following a descriptive norm, as seen in the bystander effect (Darley & Latane, 1968).

In applying regret aversion to our results, this theory makes no distinction between ingroups and outgroups and thus, naturally assumes that people will follow the overall norm (which is what we found) rather than actively avoiding conforming to the outgroup norm. In other words, responsibility can be diffused just as easily to outgroup members as it can to ingroup members. Furthermore, regret aversion was designed to explain irrational biases towards inaction (Baron & Ritov, 1994; Kahneman & Miller, 1986). Just like it does not require that a default action offers useful information, it would not assume that descriptive norms offer useful information. Thus, it can easily account for our finding that people conform to arbitrary norms that offer no useful information.

Another reason why people may tend to conform to the overall norm is that they view it as the status quo. Status quo biases, such as the endowment effect, involve a preference towards the current state of affairs (Samuelson & Zeckhauser, 1988; Thaler, 1980). It seems reasonable that people would treat the overall descriptive norm as a societal status quo.

The most prominent explanation for these status quo biases is prospect theory (Kahneman & Tversky, 1979; Kahneman, Knetsch, & Thaler, 1991). If people treat the overall descriptive norm as a reference point, then they may tend to evaluate all other options relative to it. Any way in which an option is better than the overall descriptive norm will be seen as a gain while any way in which an option is worse than the overall descriptive norm will be seen as a loss. The notion of loss aversion then suggests that these losses loom larger than gains, such that shifting away from the overall descriptive norm tends to have a net negative value. However, the moral dilemmas presented in the current paper were hypothetical and, even hypothetically, did not offer any clear personal gains or losses to the participant. Thus, unless loss aversion extends beyond real, personal outcomes, it may not be able to explain our particular findings.

Pre-decisional information distortion is another general mechanism we discussed in Chapter 2 that could potentially explain the descriptive norm effect. Pre-decisional information distortion proposes that we tend to interpret and think about features of the options in a way that is favourable to the target option (Chaxel et al., 2018; Russo et al., 1996; Carlson et al., 2006). For example, we may be biased to initially think about reasons to conform to a descriptive norm. This may then hinder our ability to think about reasons to not follow the descriptive norm (Johnson et al., 2007), resulting in preference for the descriptive norm. As with the mechanisms discussed above, pre-decisional information distortion does not distinguish between ingroups and outgroups, nor between arbitrary and non-arbitrary norms, allowing it to explain our findings.

As some support for the role of pre-decisional information distortion in the descriptive norm effect, Melnyk et al. (2011) found that participants were more likely to list positive aspects than negative aspects of conforming to a descriptive norm, especially when participants were encouraged to deliberate on a decision. However, Melnyk et al. (2011) did not counterbalance the descriptive norm, meaning the descriptive norm always favoured the same behaviour (specifically buying environmentally-friendly potatoes) when presented. Instead, they compared the effect of a descriptive norm that environmentally-friendly potatoes were popular to an injunctive norm recommending environmentally-friendly potatoes. It is therefore possible that the descriptive norm had no impact on participant's positive thoughts about potatoes, with the injunctive norm instead inhibiting positive thoughts. For example, Melnyk et al. (2011) suggest that people may feel that injunctive norms restrict their freedom and thus, may tend to come up with more reasons to avoid conforming to them (see reactance theory Brehm & Brehm, 2013). Though not strong evidence that pre-decisional information distortion underlies the descriptive norm effect, Melnyk et al.'s (2011) results are at least consistent with this explanation.

The current study was not designed to distinguish between these alternative theories of why people may tend to adhere to the overall descriptive norm. Further work could look to test how well these general mechanisms apply to the descriptive norm effect. For example, if people follow the overall descriptive norm because it allows them to reduce their sense of guilt in the event that their action has negative consequences, we would expect the descriptive norm effect to be stronger when the decision has important consequences external to the participant, increasing the potential level of guilt. In contrast, loss aversion and anticipated regret explanations would predict the descriptive norm effect to be stronger when the decision has large personal consequences for the participant. If people follow norms due to a conditioned or evolved anxiety response to violating norms, then they should be influenced by the descriptive norm just as much, regardless of the above changes, provided anxiety is controlled for. All of these alternative explanations we have offered suggest that negative thoughts and emotions such as guilt, regret and anxiety are increased when deviating from a norm. Self-report and physiological tests could look at whether this is the case or whether entirely different explanations are needed.

Though this paper is theoretically driven, the fact that descriptive norms are commonly used in real-world behavioural interventions encourages consideration of the practical implications of our findings. The obvious implication of our results is that people may tend to conform to popular behaviour, even when that behaviour is predominantly engaged in by outgroup members. For example, if a particular social group tends to engage in an undesirable, polarised behaviour, our findings suggest that increasing exposure of these individuals to people outside their group is likely to decrease the incidence of their undesirable behaviour. In contrast, self-categorisation theory predicts that increased exposure to outgroups would encourage such individuals to further polarise towards their socially undesirable behaviour.

In fact, our results suggest that interactions with heterogeneous groups, comprised of people with different opinions, might lead individuals to arrive at less, rather than more, polarised decisions. This is in line with findings that more diverse groups tend to be more productive (Hoffman & Maier, 1961) and that interaction between groups can reduce intergroup prejudice (Allport, 1954; Pettigrew & Tropp, 2006). By supporting the idea that people care about overall norms, our findings suggest that people may have vastly different responses to outgroup information than would be assumed if relying solely on self-categorisation theory.

In this paper we found that people's preference shifted towards an option that was popular (the descriptive norm), even when that option was popular amongst outgroup members. This occurred even when the outgroup was based on strongly divisive issues such as political affiliation or important social issues. Prior to this research, it would have been intuitive to predict that people would actively avoid following the norms of outgroups they are clearly opposed to (such as an opposing political party). This prediction is made by self-categorisation theory, a particularly prominent explanation of the descriptive norm effect. Thus, self-categorisation theory cannot explain our results. Instead, other theoretical explanations of the descriptive norm effect are needed that assume people have a more general desire to conform towards norms, even when they come from an outgroup. We have considered how a number of more general mechanisms drawn from broader findings on conditioned emotional responses, anticipated regret, reference point effects and pre-decisional information distortion could potentially explain our findings.

Chapter 9

General Discussion

Across 13 experiments, this thesis looked at the mechanisms underlying two heavily-studied choice biases. The first effect (the endowment effect) has maintained substantial research interest due to its implications for how we make choices. In contrast, the second effect (the descriptive norm effect) had received little theoretical attention. Instead, interest and research around the descriptive norm effect has generally been focused on its practical implications. Through experimentally testing theoretical explanations for these effects, we found that a prominent theory of the endowment effect (loss aversion) remained a plausible explanation despite previous evidence to the contrary. Meanwhile, none of the prominent explanations of the descriptive norm effect were able to fully account for our findings.

9.1 Understanding the descriptive norm effect

Theoretical interest in choice biases has been heavily driven by the fact that many choice biases violate axioms of rational choice theories. The endowment effect, for example, violates the axiom of reference independence, which posits that decision makers should care about the final outcome of their decision independent of their starting state (Thaler, 1980). This indicates that other mechanisms are driving choices. Studying choice biases offers interesting insights into what these other mechanisms are. A corollary of this is that choices that can be reasonably explained and do not violate such axioms are potentially less theoretically interesting.

In the case of the descriptive norm effect, two straight forward explanations have been commonly applied to explain this effect based on objective reasons to conform to the descriptive norm. The informational account proposed that if an option is popular amongst other people, it is likely because that option is indeed better (Cialdini et al., 1990). The social sanction account proposed that people conform to descriptive norms in an effort to minimise negative social sanctions and maximise positive ones (Ajzen & Fishbein, 1980; Bendor & Swistak, 2001). Given the existence of these objective reasons to conform to descriptive norms, it is perhaps unsurprising that experimental tests of the mechanisms driving the descriptive norm effect were rare. This thesis provided new insights into why people show descriptive norm effects.

9.1.1 Arbitrary descriptive norms

In contrast to the prominent, objectivity-based explanations of the descriptive norm effect, Chapter 6 found that people continued to conform towards a descriptive norm even when this norm was entirely arbitrary. These arbitrary norms occurred due to random allocation and thus, provided no useful information about what other people considered to be better and/or more socially acceptable. The fact that our participants' decisions were private further emphasises that our findings of conformity to arbitrary norms cannot be attributed to a fear of social sanctions. This study established that the descriptive norm effect occurs even when there is no objective reason why it should occur.

If the descriptive norm effect cannot simply be attributed to objective reasons why people should follow descriptive norms, this suggests the descriptive norm effect may be more theoretically interesting than previously thought. It is our hope that this result will spur further interest in the mechanisms actually underlying the descriptive norm effect.

Better understanding the mechanisms underlying norm effects is not only important from the perspective of trying to better understand and model human decision making but also for being better able to predict when people will conform to descriptive norms. A substantial amount of research has looked at influencing important, real-world choices and behaviour through descriptive norms as well as looking at the moderating factors that increase the strength of these effects. Having a stronger understanding of why people follow descriptive norms could offer important guidance for these more practically-driven studies.

On the topic of more applied studies, a notable limitation of our arbitrary norm experiments is revealed through performance on the comprehension check. A relatively high portion of participants failed the comprehension check in our arbitrary norm experiments, indicating that the task was quite hard to understand. Our choice scenario was quite contrived in order to control for a range of potential confounds. This opens up the possibility that participants could have been overwhelmed by the details of the task, decreasing their ability to think about the decision. This could have altered their decision-making process from how they would decide during a more familiar task.

While these controlled scenarios are important for theory-testing, it would be worthwhile seeing if conformity to arbitrary norms can be observed in more familiar, realistic settings. For example, an applied study could look at whether people tend to follow social norms that have arisen due to previously-existing defaults, such as a previous default industry pension/superannuation fund. These norms would still be arbitrary, but the real-world context could make the decision less overwhelming. This would have the added benefit of adding real-world incentives that were lacking from our experiments.

9.1.2 Outgroup descriptive norms

To better understand why people were following descriptive norms, we then turned to testing another prominent account of the descriptive norm effect, self-categorisation theory. Self-categorisation theory proposes that an individual will follow ingroup descriptive norms, and avoid following outgroup descriptive norms, in an effort to maintain a personal sense of identity with their ingroup (Turner et al., 1987; Hogg et al., 1990). Crucially, as long as a behaviour is associated with a salient identity, self-categorisation theory would expect people to follow that norm, even if the existence of such a norm has occurred through arbitrary reasons. Consequently, self-categorisation theory can account for our observed effect of arbitrary descriptive norms.

Counter to the predictions of self-categorisation theory, Chapter 8 found that participants did not avoid conforming towards outgroup descriptive norms. We presented participants with an outgroup descriptive norm that favoured the opposite action to an ingroup descriptive norm. Rather than participants shifting away from this outgroup descriptive norm, further biasing them towards the ingroup descriptive norm, we found that participants' preferences tended to shift towards the outgroup descriptive norm when it was presented. This occurred even when defining ingroups and outgroups based on salient social and political issues that should represent important aspects of people's personal identity (Greene, 1999). While self-categorisation may still contribute to the descriptive norm effect under other conditions, our findings suggest that self-categorisation cannot fully account for the descriptive norm effect, even under conditions that were intended to be conducive to self-categorisation mechanisms.

When combined with our finding that people conform to arbitrary descriptive norms, these results indicate that none of the three most prominent explanations of the descriptive norm effect (the informational, social sanction and self-categorisation theories) can fully account for people's bias to conform to descriptive norms. Instead, our findings indicate that people tend to conform towards descriptive norms even when they come from an outgroup and/or do not reflect people's actual preferences.

These results represent an important step towards understanding descriptive norms. However, one limitation of our studies is that they only falsify theories. We have not established a specific alternative theory as a superior account of the descriptive norm effect. Turning to the literature on other choice biases, we can find a number of cognitive mechanisms that could be generalised to explain our descriptive norm effect findings.

9.1.3 New, speculative explanations of the descriptive norm effect

In Chapter 2, we identified a number of mechanisms from across the choice bias literature that can potentially explain a wide range of choice biases. Having established that none of the prominent descriptive norm accounts can explain our results, this opens up the question of whether one of these more generalisable mechanisms is driving the descriptive norm effect. Thus, it is worth considering how well our results could be explained by generalisable mechanisms such as loss aversion, regret aversion, pre-decisional information distortion, item attention and attribute weighting. We leave testing of these explanations to future work, though offer some potential avenues through which these mechanisms could be tested.

Loss aversion

In explaining the descriptive norm effect, loss aversion would simply assume that we treat descriptive norms as a reference point, favouring the norm so as to avoid losses relative to it (Tversky & Kahneman, 1991). Under loss aversion, this seems like a reasonable assumption given descriptive norms essentially represent a group-level status quo and thus, could be functionally equivalent to other status quo biases. In fact, this idea was mentioned off-hand by Kahneman (1992), although they focused more on norms in terms of historical tendencies that the decision maker can easily recall (e.g. the typical price of a movie ticket) rather than social norms. Nevertheless, we are not aware of research testing loss aversion as an explanation of the descriptive norm effect.

Our observed effect of arbitrary descriptive norms is easily accounted for by loss aversion, given loss aversion was originally proposed to explain biases that also had no objective reason why they should occur. In essence, all that matters is that people's reference point shifts towards the popular option, regardless of the reasons for this popularity. Furthermore, ingroup and outgroup behaviour are both relevant to the overall status quo, and so loss aversion would also predict our finding that participants' preferences shifted towards the overall descriptive norm, rather than shifting away from the outgroup descriptive norm.

With that said, the choice stimuli in our descriptive norm experiments were hypothetical and did not involve any explicit personal consequences for participants. It is therefore unclear whether participants would have actually experienced any sense of potential loss during our experiments. Future work could better establish whether evidence of loss aversion extends to the descriptive norm effect. For example, if a descriptive norm is set as a reference point, we should expect options with larger gains and losses relative to this reference point to be evaluated more negatively than options that are more similar to the descriptive norm. Additionally, increasing the personal consequences of a decision should increase potential losses and thus, strengthen the descriptive norm effect according to loss aversion.

Regret aversion

From a regret aversion perspective, it is quite possible that people consider deviating from a descriptive norm to involve taking on greater responsibility for their actions. Following a descriptive norm may allow the decision makers to offset responsibility for their actions onto all the other people who also tended to act that way. Thus, deviating from a norm could increase the decision maker's involvement with a decision, and thus their potential for regret (Kahneman & Miller, 1986). Given people tend to be averse to experiencing regret (Baron & Ritov, 1994), this could then explain the tendency to conform towards descriptive norms. In essence, we may want to follow social norms so that we do not feel responsible if the decision turns out poorly.

Offloading of the responsibility (and thus potential regret) of a decision would presumably occur regardless of whether this offloading is onto ingroup or outgroup members. Thus, regret aversion can explain our finding that people conform towards, rather than away from, outgroup descriptive norms.

It is arguable whether the effect of arbitrary norms are well accounted for by regret aversion. It could be that people experience less regret when they know that other people chose to act the same way, indicating that other people made the same mistake. Under this assumption, norms might only be expected to affect regret when they reflect people's actual choices. However, Steffel and Williams (2018) found that people even experienced less regret when offloading a decision to an inexperienced decision maker that would not make an informed decision. This finding suggests that the quality of other people's choices may not be important when offloading responsibility to them. Conforming to arbitrary norms may be analogous to offloading responsibility to inexperienced, uninformed decision makers; we are willing to offload responsibility to others even if the outcome will then be uninformed or arbitrarily random.

Both loss aversion and regret aversion suggest that conformity to descriptive norms should reduce the experience of aversive reactions (loss or regret) when conforming to the norm. This could be tested either through self-report ratings or through physiological or neural measures of aversive arousal stemming from decisions. For example, regret aversion indicates that people should rate and exhibit greater signs of regret when a negative outcome results from violating a norm compared to conforming to a norm. Establishing whether such an effect occurs would be an important step in determining whether regret aversion may underly the descriptive norm effect.

Pre-decisional information distortion

Pre-decisional information distortion offers another potential explanation of our descriptive norm effect findings. Specifically, being informed of a descriptive norm may cause people to interpret and think about the available information in such a way that it favours the popular option. Query theory, for example, could suggest that we tend to initially focus on reasons to conform towards a descriptive norm, even if it is arbitrary (Johnson et al., 2007). This may then subsequently inhibit our ability to think about reasons to favour any less popular options (Johnson et al., 2007). Consistent with this, computational modelling of social norm effects has found that a parameter representing information processing (drift rate), rather than shifts in a decision threshold parameter, tends to be associated with social norm effects in visual perception decisions (Germar et al., 2014; Germar et al., 2016; Germar & Mojzisch, 2019). This finding is consistent with, though not exclusively evidence for, pre-decisional information distortion playing a role in descriptive norm effects.

In the case of our studies, informing participants that the descriptive norm was to report the robber during our "robin hood" dilemma may have encouraged participants to initially think about the importance of laws in functioning societies, the potential dangers of allowing vigilante resource distribution etc. In contrast, stating that the norm was to not report the robber may have induced focus on the excessive wealth banks possess, the fact that insurance could cover the bank's losses, the importance of helping the downtrodden etc. According to query theory, even if a norm is arbitrary it should still induce pre-decisional information distortion as long as the decision maker first thinks about reason's to conform.

It is less clear how well pre-decisional information distortion would account for our finding that people conform towards outgroup descriptive norms. Intuitively, we might assume that people would be biased towards thinking about reasons to not follow an outgroup norm. For example, a Democrat may want to come up with reasons to not follow a Republican norm and vice versa. In contrast, it could be that people are biased to think about reasons favouring any descriptive norm, whether it is from an ingroup or outgroup, in an effort to justify the popularity of that option (Willemsen et al., 2011). For example, people may have some meta-cognitive goal of consistency, where they try to simplify the world by mentally having all information support the same option (Simon et al., 2004; Festinger, 1962). Our results suggest that, if pre-decisional information distortion is to explain the descriptive norm effect, the latter assumption must hold.

Future work could test this by having participants list their thoughts relating to each option. We would expect that participants will be biased towards listing more reasons to conform towards the overall descriptive norm (Melnyk et al., 2011). Additionally, if neutral attributes were presented, we should expect participants to rate them as more strongly favouring the overall descriptive norm (Carlson et al., 2006). If it were found that people tended to distort information against an outgroup descriptive norm or that descriptive norms had no impact on how people evaluated information, this would suggest that pre-decisional information distortion cannot explain our findings.

Item attention and attribute weighting

Finally, it could be that descriptive norms increase the salience of the popular option. According to item attention mechanisms, increased attention towards a popular option could amplify the perceived valence of that option. This then results in accumulating stronger preferences in relation to the popular option (Shimojo et al., 2003; Krajbich, Lu, Camerer, & Rangel, 2012). From an attribute weighting perspective, salience theory (Bordalo et al., 2012a) and the associative accumulation model (Bhatia, 2013) suggest that the descriptive norm increases activation (and thus the weighting) of the attributes that the salient, popular option scores highly on. Increased weight being placed on attributes that a popular option scores well on, would then lead to higher perceived value of the popular option, explaining the descriptive norm effect.

These attentional processes make no assumptions about the meaningfulness of a norm. Instead, it is well established that attentional biases can occur automatically, even towards irrelevant stimuli (Hommel, Pratt, Colzato, & Godijn, 2001; Ranzini, Dehaene, Piazza, & Hubbard, 2009; Gaspelin & Luck, 2018). Thus, item attention and attribute weighting naturally account for our finding that people continued to conform towards descriptive norms that were arbitrary.

Additionally, while ingroup descriptive norms may be more salient than outgroup descriptive norms, outgroup descriptive norms should still attract some attention. Any shift of attention towards the option popular amongst the outgroup would then be expected to shift preferences towards this option.

Given the emphasis placed on attention, eye-tracking studies could offer strong tests of these theories. Future work could test whether increased attention (as indicated by longer gaze duration or more fixations) is given to a popular option (item attention) or towards attributes that the popular option scores highly on (attribute weighting). With that said, item attention and attribute weighting mechanisms predict that the standard descriptive norm effect should reverse when choosing between undesirable options. Our study of undesirable choices has offered some evidence against this, as discussed shortly.

9.2 Undesirable choices

Studying undesirable choices is interesting because theories from across the choice bias literature make quite strong predictions as to how choice biases should change based on the desirability of the choice options. The majority of prominent theories predict that choice biases will occur in the same direction regardless of whether the choice options are desirable or undesirable (see Appendix A on page 143 for a review). A particularly important example of this is loss aversion, where emphasis is only placed on the relative value of options to a reference point, such that the absolute desirability should have no impact on choice biases (Kahneman & Tversky, 1979; Tversky & Kahneman, 1991). Thus, previous findings that some choice biases may reverse when choosing between undesirable options was particularly interesting to us (Brenner et al., 2007; Armel et al., 2008; Houston et al., 1989).

9.2.1 What do we now know about undesirable choices?

In an effort to see whether multiple choice bias reverse when choosing between undesirable options, we ran a pilot study that tested a range of prominent choice biases on both desirable and undesirable choice scenarios. Interestingly, we did not find that any choice biases showed signs of reversals. While some manipulations failed to show a significant effect, for all choice biases that did show a significant effect, none of them reversed. Specifically, we found that participants continued to show the typical attraction, phantom decoy, descriptive norm, injunctive norm, default biases and endowment effects even for undesirable choices.

This pilot study represents initial evidence that multiple choice biases do not show reversals for undesirable choices. It would be interesting for further research to better establish these findings, seeing how well they generalise. However, what we found particularly interesting was the fact that the endowment effect did not reverse. This finding contradicts previous findings by Brenner et al. (2007) and Dertwinkel-Kalt and Köhler (2016), who found that participants wanted to switch away from an undesirable option more when they were hypothetically endowed ownership of it. Given the theoretical implications of choice biases reversing for undesirable choices, determining why our results contradicted those of Brenner et al. (2007) and Dertwinkel-Kalt and Köhler (2016) was particularly important to us.

In Chapter 4 we found that we could indeed observe reversals of the endowment effect for undesirable choices. Importantly, we established for the first time that these reversals could occur even for choices with real consequences. However, what we found was that we could only observe these reversals when a participant's ability to compare the available options was limited. This was done either by explicitly hiding information about the non-endowed alternative or by enforcing a timed response deadline.

This was consistent with Dertwinkel-Kalt and Köhler's (2016) proposal that hypothetical (rather than incentivised) choices are more likely to induce endowment reversals, because they do not sufficiently motivate people to compare the available options. However, we failed to find an endowment reversal even when using hypothetical choices in our pilot study.

Our findings suggest that the difference between our pilot study's results and Dertwinkel-Kalt and Köhler's (2016) may be because our stimuli made it easier for participants to compare the options. Whereas Brenner et al. (2007) and Dertwinkel-Kalt and Köhler (2016) described their options in paragraphs of text, our pilot study presented options in tables that clearly outlined how the options differed. These tables may have made it easier to directly compare the options, encouraging comparisons even when the decisions were hypothetical (Noguchi & Stewart, 2014; Kelly, 1993; Smerecnik et al., 2010).

The question was then, why does the endowment effect reverse for undesirable choices when comparisons are limited, but not when they are allowed and encouraged? We found that a decision maker's prior reference point and expectations may drive these results. Prior to most experimental studies of choice biases, participants would not have much context for the choice and thus, should have a relatively neutral reference point. An undesirable endowment is then worse than this neutral reference point and thus, if not comparing it to any alternatives, will tend to be avoided. Consistent with this, we found that presenting practice gambles that were superior to the subsequently available gambles (intended to heighten a participant's expectations) resulted in reversals of the endowment effect. These reversals occurred not only for undesirable choice stimuli but also for desirable choice stimuli. In contrast, when participants were presented with an inferior practice gamble (lowering their expectations), participants tended to favour the endowed option regardless of whether it was desirable or undesirable. These results indicate that limited comparisons and expectations may be crucial to the endowment effect reversing, rather than the absolute desirability of the options. This finding alters the theoretical implications of endowment reversals.

9.2.2 Theoretical implications of our undesirable choice findings

The fact that we could show endowment reversals even for desirable stimuli (when expectations were high) means that endowment reversals are not simply related to the desirability of options. This suggests that these endowment reversals may not be testing theories of choice biases in the manner intended. We have reviewed how various theories make strong predictions as to whether choice biases will reverse for undesirable options (see Appendix A on page 143). Our results suggest that reversals instead occur regardless of desirability, based on separate mechanisms related to people's expectations.

Consistent with this, we showed that our results could be accounted for by loss aversion if we adapt prospect theory slightly. Specifically, we said that a decision maker's reference point is not only influenced by their current state (as is often assumed by prospect theory) but also by their previous state. This explains the typical endowment effect, when comparisons are allowed and encouraged, based on the reference point being closer to the endowed option than the competitor, such that the endowed option has fewer perceived losses. Meanwhile, when comparisons are limited, the endowed option is favoured if it is better than one's reference point, otherwise it is avoided. Indeed, we showed that such a model was able to qualitatively and quantitatively account for our data.

Interestingly, our results actually represent evidence against theories that predict an effect of desirability on choice biases. Like Dertwinkel-Kalt and Köhler (2016), when we allowed participants to freely compare the options and incentivised them to do so, the endowment effect did not reverse. Similarly, when we presented a range of undesirable choices in the pilot study in Chapter 3, we failed to find reversals for any of the studied choice biases.

These findings are inconsistent with item attention and attribute weighting explanations of these effects, which assume that the valence of a target option is amplified. This amplification predicts that undesirable targets should appear more undesirable than competitors. Thus, these theories cannot explain our findings that various choice biases remained consistent regardless of desirability under conditions that encouraged comparisons between the options. Indeed, we discussed earlier how item attention or attribute weighting mechanisms could explain our findings that people conform to descriptive norms even when they are arbitrary or come from outgroups. The fact that our pilot study found that the descriptive and injunctive norm effects did not reverse for undesirable choice stimuli offers initial evidence against these theories.

As it stands, the choice bias literature is yet to establish whether different choice biases are meaningfully distinct or instead could be driven by the same mechanisms. Is the endowment effect just a variant of the mere exposure effect (Plott & Zeiler, 2007; Tom et al., 2007)? Are social norm effects really just an example of status quo biases? Are all of these biases being influenced by a common set of cognitive mechanisms? The fact that we found that multiple biases did not reverse for undesirable choices leaves open the possibility that each of these biases are driven by the same underlying mechanisms. Future work should look further at whether these biases might be functionally dissociated in some other way.

9.3 Practical implications

9.3.1 Comparing biases

One factor hindering our ability to integrate knowledge across choice biases is the fact that choice biases are frequently studied under quite distinct conditions. For example, decoy effects such as the attraction effect, tend to be studied using controlled stimuli that vary along two quantitatively described attributes (Huber et al., 1982; Pettibone & Wedell, 2000; Simonson, 1989). In stark contrast, we have discussed how social norm effects are generally studied using more realistic, but less controlled choice stimuli (Cialdini et al., 1990; Passafaro et al., 2019; Rimal & Real, 2003; Vinnell et al., 2019). Endowment effects are generally studied using incentivised, but more trivial choices relating to physically-present objects (Knetsch, 1989; Kahneman et al., 1990). Meanwhile, mere exposure effects almost universally look at preferences between experiential stimuli such as images (Zajonc, 1968; Johnson et al., 2007; Newell & Shanks, 2007; Stafford & Grimes, 2012; Tom et al., 2007). This makes it difficult to compare findings across these choice biases.

Any effort to compare all of these choice biases under a standard set of conditions would require deviating from the typical way in which some of these choice biases are studied. The risk that any effect will not generalise to such conditions may limit willingness to more directly compare choice biases. Without being able to directly compare choice biases, it will be difficult to establish whether a common set of general mechanisms might explain multiple biases. Thus, a positive outcome from our pilot study that looked at multiple choice biases was establishing that at least some of these choice biases can be observed under a standard set of conditions.

We showed that the attraction, phantom decoy, endowment, default, descriptive norm and injunctive norm effects could be observed using options that varied along controlled attributes, similar to how decoy effects are typically studied. This finding may help future work look at directly contrasting these biases further. Future efforts to functionally dissociate these biases could potentially study each of these biases at once, rather than restricting studies of each bias to the typical conditions in which it has previously been studied. With that said, we failed to observe significant scarcity, mere exposure or leader-drive primacy effects under these conditions. Further exploration into whether these biases can occur under the same conditions as other biases could aid our ability to directly compare these biases. On the other hand, establishing that these biases occur under distinct conditions to other choice biases may indicate that they rely on meaningfully different mechanisms.

9.3.2 Limited processing

A generally important reminder from our research into undesirable choice stimuli is that the impact of manipulations on choices can differ depending on the extent to which people are processing and comparing the available choice options. At the extreme, we found that the impact of endowing an option completely reversed depending on the extent to which options were compared. This emphasises the importance of considering how alternatives are likely to be processed both when studying choice biases as well as when trying to apply choice biases to influence real-world decisions. For example, research may find that a marketing message is likely to be persuasive during initial, controlled tests. However, this message could have a fundamentally different impact when it is subsequently presented in a noisy, real-world environment, full of distractions that limit people's ability to fully process the information.

In light of this, it is worth considering the possibility that participants were not fully processing the available information in our own studies of the descriptive norm effect. In Chapter 6 and Chapter 8, we presented participants with hypothetical choices without any explicit incentives for participants to engage with the choice. Thus, it is possible that our participants did not fully process the available information and made decisions based on a limited set of information in non-obvious ways. For example, it may have been that the norm statements stood out more than the actual moral dilemma, because our participants were expecting see the moral dilemma but not the norm statement. This may have meant that the presented norms were not simply influencing how participants felt while evaluating the moral dilemma.

Instead, participants may have only focused on following the norm, independent of how they actually felt about the moral dilemma. While this still would reflect conformity to descriptive norms, it potentially may mean that our results would not generalise to conditions in which participants are heavily engaged with the decision. Determining how well our descriptive norm findings generalise to more explicitly incentivised settings is an area of ongoing research.

With that said, we did make attempts to account for how well participants were processing the available information in our descriptive norm experiments. This included the obvious approach of including understanding checks that allowed us to exclude participants who failed to engage with and understand the instructions. Additionally, we ran control experiments to account for some of the most likely ways in which our results could have been attributed to limited information processing. If participants were purely focusing on whichever option was mentioned in the descriptive norm statement, independent of its popularity (e.g. some kind of exposure effect), then they would not be sensitive to the popularity of the options. Counter to this, Experiment 3 of Chapter 6 found that participants conformed towards the popular option even when they were only informed as to which option was unpopular. Furthermore, Experiment 4 of Chapter 6 found that participants were no longer affected by descriptive norms when they were informed that the descriptive norm related to a different dilemma. This indicated that participants were sensitive to the relevance of the norm to the dilemma and thus, were indeed engaging with the presented information.

9.3.3 Initial expectations

In Chapter 4 we found that preferences towards an endowed option could be entirely reversed based on whether a superior or inferior initial reference point was set. This result suggests that pre-existing expectations can have a drastic impact on how people respond to endowments and potentially a range of other choice interventions. For example, a common marketing technique is to offer samples/trials of products, potentially creating an endowment effect through endowing ownership of the product. Our results suggest that this could become particularly problematic if initial marketing has overstated the benefits of the product. If the customer has particularly high expectations of sampled product that are then not met, they may become encouraged to switch to a competing product. This is essentially akin to the endowment reversal we observed when a superior reference point was set. Even if the sampled product is desirable and would otherwise be enjoyed by the consumer, the high expectations and endowment of the sample may cause them to move to a competitor.

9.3.4 Norm-based interventions

Given the extensive use of choice biases in efforts to nudge people towards more pro-social behaviour (Lehner et al., 2016; Tannenbaum, Fox, & Rogers, 2017; Voyer, 2015), better understanding these biases has important practical implications. The better we understand choice biases, the better we will be able to predict when they will affect people's choices. Thus, better theories could help adjust nudging efforts to maximise their impact. This is especially the case for the descriptive norm effect. Not only have descriptive norms had relatively little empirical theory-testing, but they are heavily utilised both in corporate marketing efforts (Chang, Yu, & Lu, 2015) and efforts to nudge people towards more pro-social decisions (Alm et al., 2019; Behavioural Insights Team, 2013; Czajkowski et al., 2019, e101110; Ecker et al., 2019; Wenzel, 2019).

Our finding that people conform towards even arbitrary norms suggests that descriptive norms may have an even broader impact than previously thought. If people are willing to follow arbitrary norms then the cause of an option's popularity may not always be important. This offers the possibility that decision makers will be influenced by a descriptive norm even when they know that the norm is not indicative of other people's preferences. Such arbitrary norms could occur when a large number of people do not care strongly about a decision and so simply make choices based on some other bias, such as sticking with a default (Johnson & Goldstein, 2003; Choi et al., 2005). In such a case, the lack of interest in the decision means that other people's choices carry little information and are not indicative of potential social sanctions. Nevertheless, our results suggest that the descriptive norm effect may have an exacerbating effect, further pushing people towards the default option even if it stops being a default option. What this suggests is that establishing an option as "popular" in any way, regardless of people's preferences, may have a lasting impact on other people's preferences. It would be interesting for future work to see if the impact of arbitrary norms extends into more important life decisions. For example, are people more likely to choose a popular pension fund even if they know that its popularity is only because it was previously a default option (Choi et al., 2005).

The broad impact of descriptive norms were again emphasised in our finding that participants even conformed towards outgroup descriptive norms. This finding indicates that knowing an option is popular may be sufficient to incentivise a decision-maker to conform, even if they do not identify with the people the option is popular amongst. A consequence of this is that accounting for identity in descriptive norm nudging efforts may be less important than might otherwise be expected.

Previously, self-categorisation theory would have suggested that targeted marketing is especially important when presenting norm-based appeals, given the risk that people would be polarised away from the norms of a group they do not identify with. For example, telling a smoker that smoking is unpopular among some social group could potentially further entrench their smoking behaviour if they do not identify with that social group. Our results suggest this may generally not be the case. Instead, we suggest that it might be possible to mitigate undesirable behaviour within a particular social group by informing them that members of other social groups do not engage in the behaviour. On the flip side, this could also explain why people conform to anti-social descriptive norms such as corruption, even if they are unlikely to initially identify with corrupt people (Abbink et al., 2018; Bicchieri & Xiao, 2009; Köbis et al., 2015).

While we found that participants conformed towards, rather than away from outgroup descriptive norms, it is worth noting that this does not mean that group identity is irrelevant to the effect of descriptive norms. In fact, Experiment 4 in Chapter 6 found that participants were more heavily influenced by ingroup descriptive norms than outgroup descriptive norms. Instead, it seems that the effects of ingroup and outgroup descriptive norms may differ in degree, but not direction.

9.4 Concluding remarks

Subjective, preferential choices are a pervasive aspect of our lives. They determine governance in democracies, underlie demand in economic systems and impact our heath, wellbeing and lifestyle. Our choices are not perfect though. The existence of choice biases means that all of these aspects of our lives can be influenced in often subtle ways. Pervasive efforts to manipulate people's choices, ranging from marketing efforts to increase the popularity of a product through to political groups trying to influence democratic votes, underscore the importance of having a strong, public understanding of how people make choices. Choice biases offer a unique insight into the mechanisms driving choices; their study has vastly furthered our understanding of the human decision making process. However, we are still a far way off from having a broad, well-rounded, integrative theory of how people make choices.

This thesis has hoped to further our understanding of choices through studying two prominent choice biases, the endowment effect and the descriptive norm effect. Our research into the endowment effect exemplified the importance of focusing on how people are processing information and the expectations people have when making choices. We found evidence that the endowment effect could entirely reverse depending on a participant's expectations and the extent to which they were comparing the choice options. Crucially, we showed that these results could be quite naturally accounted for by prospect theory, meaning that loss aversion remains a plausible explanation of the endowment effect despite previous claims to the contrary.

In the case of the descriptive norm effect, we found that three prominent theories (the informational, social sanction and self-categorisation accounts) are unable to fully explain the descriptive norm effect. While our studies focused on testing these prominent accounts, we have not offered strong evidence in favour of a particular alternative theory. It is our hope that this research will encourage further work into the mechanisms driving the descriptive norm effect. Having shown that the descriptive norm effect is not as intuitive as previously thought, it may now pose a more interesting area for future research. We have discussed how a number of more general mechanisms could be applied to the descriptive norm effect and how they could be tested. We have taken some early, but nevertheless important steps towards understanding the pervasive and widely applied effect of descriptive norms.

It is inevitable that researchers will bring personal biases into their study of psychological phenomena. Economists may be more interested in being able to predict behaviour while policy and market researchers may focus on using choice biases to influence choices. From a cognitive psychology perspective, our bias is towards understanding the cognitive mechanisms driving choices. This thesis has offered new insights and evidence relating to why people show choice biases. For us, this understanding is interesting in and of itself. However, we hope that as this, and future work, continues to provide insights into the mechanisms driving people's choices, this understanding will aid other researchers in better predicting human behaviour and nudging people towards making better decisions.

Bibliography

Abbink, K., Freidin, E., Gangadharan, L., & Moro, R. (2018). The effect of social norms on bribe offers. *The Journal of Law, Economics, and Organization*, *34*(3), 457–474.

Abdellaoui, M., Bleichrodt, H., l'Haridon, O., & Van Dolder, D. (2016). Measuring loss aversion under ambiguity: A method to make prospect theory completely observable. *Journal of Risk and Uncertainty*, *52*(1), 1–20.

Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science*, *53*(10), 1659–1674.

Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*. Englewood Cliffs, NJ: Prentice-Hall.

Allport, G. W. (1954). *The Nature of Prejudice*. Cambridge, Mass: Addison-Wesley.

Alm, J., Schulze, W. D., Von Bose, C., & Yan, J. (2019). Appeals to social norms and taxpayer compliance. *Southern Economic Journal*, *86*(2), 638–666.

Anderson, C. A. (1983). Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakeable beliefs. *Journal of Experimental Social Psychology*, *19*(2), 93–108.

Ariely, D. & Wallsten, T. S. (1995). Seeking dubjective dominance in multidimensional space: An explanation of the asymmetric dominance effect. *Organizational Behavior and Human Decision Processes*, *63*(3), 223–232.

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. (1951). *Social choice and individual values* (First). New York: Wiley.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, Leadership, and Men* (pp. 177–190). Pittsburgh, PA: Carnegie Press.

Ashby, N. J., Dickert, S., & Glöckner, A. (2012). Focusing on what you own: Biased information uptake due to ownership. *Judgment and Decision Making*, *7*(3), 254–267.

Balcombe, K., Fraser, I., Williams, L., & McSorley, E. (2017). Examining the relationship between visual attention and stated preferences: A discrete choice experiment using eye-tracking. *Journal of Economic Behavior & Organization*, *144*, 238–257.

Barberis, N. & Xiong, W. (2009). What drives the disposition effect? An analysis of a long-standing preference-based explanation. *The Journal of Finance*, *64*(2), 751–784.

Baron, J. & Ritov, I. (1994). Reference Points and Omission Bias. *Organizational Behavior and Human Decision Processes*, *59*(3), 475–498.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral science*, *8*(1), 41–55.

Becker, S. W. & Brownson, F. O. (1964). What Price Ambiguity? Or the Role of Ambiguity in Decision-Making. *Journal of Political Economy*, *72*(1), 62–73.

Behavioural Insights Team. (2013). *Applying behavioural insights to organ donation: Preliminary results from a randomised controlled trial*. Cabinet Office.

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior research methods*, *43*(3), 800–813.

Belk, R. W. (1981). Determinants of consumption cue utilization in impression formation: An association derivation and experimental verification. *ACR North American Advances*, *8*, 170–175.

Bendor, J. & Swistak, P. (2001). The evolution of norms. *American Journal of Sociology*, *106*(6), 1493–1545.

Berger, J. & Heath, C. (2007). Where consumers diverge from others: Identity signaling and product domains. *Journal of Consumer Research*, *34*(2), 121–134.

Berger, J. & Heath, C. (2008). Who drives divergence? Identity signaling, outgroup dissimilarity, and the abandonment of cultural tastes. *Journal of personality and social psychology*, *95*(3), 593.

Berger, J. & Rand, L. (2008). Shifting signals to help health: Using identity-signaling to reduce risky health behaviors. *Journal of Consumer Research*, *34*(2), 121–134.

Berkowitsch, N. A., 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.

Bicchieri, C. & Xiao, E. (2009). Do the right thing: But only if others do so. *Journal of Behavioral Decision Making*, *22*(2), 191–208.

Biswas, D., Grewal, D., & Roggeveen, A. (2010). How the order of sampled experiential products affects choice. *Journal of Marketing Research*, *47*(3), 508–519.

Booij, A. S., van Praag, B. M. S., & van de Kuilen, G. (2010). A parametric analysis of prospect theory's functionals for the general population. *Theory and Decision*, *68*(1), 115–148.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2012a). Salience in experimental tests of the endowment effect. *The American Economic Review*, *102*(3), 47–52.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2012b). Salience theory of choice under risk. *Quarterly Journal of Economics*, *127*(3), 1243–1285.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, *121*(5), 803–843.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, *106*(2), 265–289.

Bornstein, R. F. & D'Agostino, P. R. (1994). The attribution and discounting of perceptual fluency: Preliminary tests of a perceptual fluency/attributional model of the mere exposure effect. *Social Cognition*, *12*(2), 103–128.

Brehm, S. S. & Brehm, J. W. (2013). *Psychological reactance: A theory of freedom and control*. New York: Academic Press.

Brenner, L., Rottenstreich, Y., Sood, S., & Bilgin, B. (2007). On the psychology of loss aversion: Possession, valence, and reversals of the endowment effect. *Journal of Consumer Research*, *34*(3), 369–376.

Breska, A., Israel, M., Maoz, K., Cohen, A., & Ben-Shakhar, G. (2011). Personally-significant information affects performance only within the focus of attention: A direct manipulation of attention. *Attention, Perception, & Psychophysics*, *73*(6), 1754–1767.

Brock, T. C. (1968). Implications of commodity theory for value change. In *Psychological foundations of attitudes* (pp. 243–275). New York: Academic Press.

Bucher, T., Collins, C., Rollo, M. E., McCaffrey, T. A., De Vlieger, N., Van der Bend, D., ... Perez-Cueto, F. J. (2016). Nudging consumers towards healthier choices: A systematic review of positional influences on food choice. *British Journal of Nutrition*, *115*(12), 2252–2263.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5.

Burger, J. M., Bell, H., Harvey, K., Johnson, J., Stewart, C., Dorian, K., & Swedroe, M. (2010). Nutritious or delicious? The effect of descriptive norm information on food choice. *Journal of Social and Clinical Psychology*, *29*(2), 228–242.

Camerer, C. & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, *5*(4), 325–370.

Carlson, K. A., Meloy, M., & Russo, J. E. (2006). Leader-driven primacy: Using attribute order to affect consumer choice. *Journal of Consumer Research*, *32*(4), 513–518.

Carmon, Z. & Ariely, D. (2000). Focusing on the forgone: How value can appear so different to buyers and sellers. *Journal of Consumer Research*, *27*(3), 360–370.

Celsi, R. L. & Olson, J. C. (1988, September 1). The role of involvement in attention and comprehension processes. *Journal of Consumer Research*, *15*(2), 210–224.

Chang, Y.-T., Yu, H., & Lu, H.-P. (2015). Persuasive messages, popularity cohesion, and message diffusion in social media marketing. *Journal of Business Research*. Special Issue on Global entrepreneurship and innovation in management, *68*(4), 777–782.

Chaxel, A.-S., Wiggins, C., & Xie, J. (2018). The impact of a limited time perspective on information distortion. *Organizational Behavior and Human Decision Processes*, *149*, 35–46.

Chen, H.-J. & Sun, T.-H. (2014). Clarifying the impact of product scarcity and perceived uniqueness in buyers' purchase behavior of games of limited-amount version. *Asia Pacific Journal of Marketing and Logistics*, *26*(2), 232–249.

Choi, J. J., Laibson, D., Madrian, B. C., & Metrick, A. (2005). Passive decisions and potent defaults. In *Analyses in the Economics of Aging* (pp. 59–78). Chicago: University of Chicago Press.

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

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., & Winter, P. L. (2006). Managing social norms for persuasive impact. *Social Influence*, *1*(1), 3–15.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026.

Coase, R. H. (1960). The problem of social cost. In *Classic papers in natural resource economics* (pp. 87–137). New York: Springer.

Cruwys, T., Platow, M. J., Angullia, S. A., Chang, J. M., Diler, S. E., Kirchner, J. L., ... Tor, V. W. (2012). Modeling of food intake is moderated by salient psychological group membership. *Appetite*, *58*(2), 754–757.

Cunningham, S. J., Turk, D. J., Macdonald, L. M., & Macrae, C. N. (2008). Yours or mine? Ownership and memory. *Consciousness and Cognition*, *17*(1), 312–318.

Czajkowski, M., Zagórska, K., & Hanley, N. (2019). Social norm nudging and preferences for household recycling. *Resource and Energy Economics*, *58*.

Darley, J. M. (1966). Fear and social comparison as determinants of conformity behavior. *Journal of Personality and Social Psychology*, *4*(1), 73–78.

Darley, J. M. & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*(4), 377–383.

De Martino, B., Camerer, C. F., & Adolphs, R. (2010). Amygdala damage eliminates monetary loss aversion. *Proceedings of the National Academy of Sciences*, *107*(8), 3788–3792.

de Bruin, W. B. & Keren, G. (2003). Order effects in sequentially judged options due to the direction of comparison. *Organizational Behavior and Human Decision Processes*, *92*(1), 91–101.

de Haan, T. & Linde, J. (2018). 'Good nudge lullaby': Choice architecture and default bias reinforcement. *The Economic Journal*, *128*(610), 1180–1206.

Dean, M., Kıbrıs, Ö., & Masatlioglu, Y. (2017). Limited attention and status quo bias. *Journal of Economic Theory*, *169*, 93–127.

Dempster, F. N. (1995). Interference and inhibition in cognition: An historical perspective. In *Interference and inhibition in cognition* (pp. 3–26). San Diego: Academic Press.

Dertwinkel-Kalt, M. & Köhler, K. (2016). Exchange asymmetries for bads? Experimental evidence. *European Economic Review*, *82*, 231–241.

Deutsch, M. & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629–636.

DeWall, C. N., Chester, D. S., & White, D. S. (2015). Can acetaminophen reduce the pain of decision-making? *Journal of Experimental Social Psychology*, *56*, 117–120.

Dhingra, N., Gorn, Z., Kener, A., & Dana, J. (2012). The default pull: An experimental demonstration of subtle default effects on preferences. *Judgment and Decision Making*, *7*(1), 69–76.

Ditto, P. H. & Jemmott, J. B. (1989). From rarity to evaluative extremity: Effects of prevalence information on evaluations of positive and negative characteristics. *Journal of Personality and Social Psychology*, *57*(1), 16–26.

Dommer, S. L. & Swaminathan, V. (2013). Explaining the endowment effect through ownership: The role of identity, gender, and self-threat. *Journal of Consumer Research*, *39*(5), 1034–1050.

Drouvelis, M. & Sonnemans, J. (2017). The endowment effect in games. *European Economic Review*, *94*, 240–262.

Ecker, A. H., Dean, K. E., Buckner, J. D., & Foster, D. W. (2019). Perceived injunctive norms and cannabis-related problems: The interactive influence of parental injunctive norms and race. *Journal of Ethnicity in Substance Abuse*, *18*(2), 211–223.

Englich, B. & Mussweiler, T. (2001). Sentencing under uncertainty: Anchoring effects in the courtroom. *Journal of Applied Social Psychology*, *31*(7), 1535–1551.

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, *32*(2), 188–200.

Eriksson, K., Strimling, P., & Coultas, J. C. (2015). Bidirectional associations between descriptive and injunctive norms. *Organizational Behavior and Human Decision Processes*, *129*, 59–69.

Farmer, J. D. & Foley, D. (2009). The economy needs agent-based modelling. *Nature*, *460*(7256), 685–686.

Farquhar, P. H. & Pratkanis, A. R. (1993). Decision structuring with phantom alternatives. *Management Science*, *39*(10), 1214–1226.

Festinger, L. (1962). *A theory of cognitive dissonance*. Stanford: Stanford university press.

Fishburn, P. C. (1970). *Utility theory for decision making*. New York: John Wiley and Sons.

Fishburn, P. C. & Kochenberger, G. A. (1979). Two-piece von Neumann-Morgenstern utility functions. *Decision Sciences*, *10*(4), 503–518.

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

Fleming, S. M., Thomas, C. L., & Dolan, R. J. (2010). Overcoming status quo bias in the human brain. *Proceedings of the National Academy of Sciences*, *107*(13), 6005–6009.

Folger, R. (1992). On wanting what we do not have. *Basic and Applied Social Psychology*, *13*(1), 123–133.

Friedman, M. & Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, *56*(4), 279–304.

Fromkin, H. L. (1970). Effects of experimentally aroused feelings of undistinctiveness upon valuation of scarce and novel experiences. *Journal of Personality and Social Psychology*, *16*(3), 521–529.

Gächter, S., Johnson, E. J., & Herrmann, A. (2007). Individual-level loss aversion in riskless and risky choices. *IZA Disussion Paper*, *2961*, 1–56.

Gal, D. & Rucker, D. D. (2018). The Loss of Loss Aversion: Will It Loom Larger Than Its Gain? *Journal of Consumer Psychology*, *28*(3), 497–516.

Gal, D. (2006). A psychological law of inertia and the illusion of loss aversion. *Judgment and Decision Making*, *1*(1), 23–32.

Galanter, E. & Pliner, P. (1974). Cross-modality matching of money against other continua. In *Sensation and measurement* (pp. 65–76). New York: Springer.

Gaspelin, N. & Luck, S. J. (2018). The role of inhibition in avoiding distraction by salient stimuli. *Trends in Cognitive Sciences*, *22*(1), 79–92.

Gawronski, B., Bodenhausen, G. V., & Becker, A. P. (2007). I like it, because I like myself: Associative self-anchoring and post-decisional change of implicit evaluations. *Journal of Experimental Social Psychology*, *43*(2), 221–232.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., ... Arnadottir, J. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, *332*(6033), 1100–1104.

Geng, S. (2016). Decision time, consideration time, and status quo bias. *Economic Inquiry*, *54*(1), 433–449.

Germar, M., Albrecht, T., Voss, A., & Mojzisch, A. (2016). Social conformity is due to biased stimulus processing: Electrophysiological and diffusion analyses. *Social Cognitive and Affective Neuroscience*, *11*(9), 1449–1459.

Germar, M. & Mojzisch, A. (2019). Learning of social norms can lead to a persistent perceptual bias: A diffusion model approach. *Journal of Experimental Social Psychology*, *84*, e103801.

Germar, M., Schlemmer, A., Krug, K., Voss, A., & Mojzisch, A. (2014). Social influence and perceptual decision making: A diffusion model analysis. *Personality and Social Psychology Bulletin*, *40*(2), 217–231.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, *35*(3), 472–482.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108.

Greene, S. (1999). Understanding Party Identification: A Social Identity Approach. *Political Psychology*, *20*(2), 393–403.

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., ... Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*, 80–97.

Gupta, P. & Harris, J. (2010). How e-WOM recommendations influence product consideration and quality of choice: A motivation to process information perspective. *Journal of Business Research*, *63*(9), 1041–1049.

Hallsworth, M., Berry, D., Sanders, M., Sallis, A., King, D., Vlaev, I., & Darzi, A. (2015). Stating appointment costs in SMS reminders reduces missed hospital appointments: Findings from two randomised controlled trials. *PLoS ONE*, *10*(9), 1–14.

Hamilton, W. D. (1971). Geometry for the selfish herd. *Journal of Theoretical Biology*, *31*(2), 295–311.

Harmon-Jones, E. & Allen, J. J. (2001). The role of affect in the mere exposure effect: Evidence from psychophysiological and individual differences approaches. *Personality and Social Psychology Bulletin*, *27*(7), 889–898.

Haslam, S. A., Oakes, P. J., McGarty, C., Turner, J. C., Reynolds, K. J., & Eggins, R. A. (1996). Stereotyping and social influence: The mediation of stereotype applicability and sharedness by the views of in-group and out-group members. *British Journal of Social Psychology*, *35*(3), 369–397.

Hauser, D. J. & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407.

Hens, T. & Vlcek, M. (2011). Does prospect theory explain the disposition effect? *Journal of Behavioral Finance*, *12*(3), 141–157.

Hergovich, A., Schott, R., & Burger, C. (2010). Biased evaluation of abstracts depending on topic and conclusion: Further evidence of a confirmation bias within scientific psychology. *Current Psychology*, *29*(3), 188–209.

Herne, K. (1998). Testing the reference-dependent model: An experiment on asymmetrically dominated reference points. *Journal of Behavioral Decision Making*, *11*(3), 181–192.

Hoffman, L. R. & Maier, N. R. (1961). Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *The Journal of Abnormal and Social Psychology*, *62*(2), 401–407.

Hogg, M. A., Turner, J. C., & Davidson, B. (1990). Polarized norms and social frames of reference: A test of the self-categorization theory of group polarization. *Basic and Applied Social Psychology*, *11*(1), 77–100.

Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic control of visual attention. *Psychological Science*, *12*(5), 360–365.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, *14*(3), 399–425.

Houston, D. A. & Sherman, S. J. (1995). Cancellation and focus: The role of shared and unique features in the choice process. *Journal of Experimental Social Psychology*, *31*(4), 357–378.

Houston, D. A., Sherman, S. J., & Baker, S. M. (1989). The influence of unique features and direction of comparison of preferences. *Journal of Experimental Social Psychology*, *25*(2), 121–141.

Huang, Y.-F. & Hsieh, P.-J. (2013). The mere exposure effect is modulated by selective attention but not visual awareness. *Vision Research*, *91*, 56–61.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 90–98.

Huff, C. & Tingley, D. (2015). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, *2*(3), 1–12.

Huh, Y. E., Vosgerau, J., & Morewedge, C. K. (2014). Social defaults: Observed choices become choice defaults. *Journal of Consumer Research*, *41*(3), 746–760.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving*, 391–422.

Johnson, E. J. & Goldstein, D. G. (2003). Do defaults save lives? *Science*, *302*, 1338–1339.

Johnson, E. J., Häubl, G., & Keinan, A. (2007). Aspects of endowment: A query theory of value construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 461–474.

Johnson, E. J., Hershey, J., Meszaros, J., & Kunreuther, H. (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty*, *7*(1), 35–51.

Kahn, I., Yeshurun, Y., Rotshtein, P., Fried, I., Ben-Bashat, D., & Hendler, T. (2002). The role of the amygdala in signaling prospective outcome of choice. *Neuron*, *33*(6), 983–994.

Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes*, *51*(2), 296–312.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, *98*(6), 1325–1348.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives*, *5*(1), 193–206.

Kahneman, D. & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*(2), 136–153.

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292.

Kahneman, D. & Tversky, A. (1984). Choices, values, and frames. *American psychologist*, *39*(4), 341–350.

Kelly, J. D. (1993). The effects of display format and data density on time spent reading statistics in text, tables and graphs. *Journalism & Mass Communication Quarterly*, *70*(1), 140–149.

Keren, G. & Gerritsen, L. E. M. (1999). On the robustness and possible accounts of ambiguity aversion. *Acta Psychologica*, *103*(1), 149–172.

Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 453–456). ACM.

Knetsch, J. L. (1989). The endowment effect and evidence of nonreversible indifference curves. *The American Economic review*, *79*(5), 1277–1284.

Köbis, N. C., Van Prooijen, J.-W., Righetti, F., & Van Lange, P. A. (2015). "Who doesn't?"—The impact of descriptive norms on corruption. *PLoS ONE*, *10*(6), e0131830.

Koop, G. J. & Johnson, J. G. (2012). The use of multiple reference points in risky decision making. *Journal of Behavioral Decision Making*, *25*(1), 49–62.

Kőszegi, B. & Rabin, M. (2006). A Model of reference-dependent preferences. *The Quarterly Journal of Economics*, *121*(4), 1133–1165.

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., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in psychology*, *3*, 1–18.

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*, *108*(33), 13852–13857.

Krizan, Z. & Baron, R. S. (2007). Group polarization and choice-dilemmas: How important is self-categorization? *European Journal of Social Psychology*, *37*(1), 191–201.

Lapinski, M. K. & Rimal, R. N. (2005). An Explication of Social Norms. *Communication Theory*, *15*(2), 127–147.

Lehner, M., Mont, O., & Heiskanen, E. (2016). Nudging–A promising tool for sustainable consumption behaviour? *Journal of Cleaner Production*, *134*, 166–177.

Levin, I. P., Huneke, M. E., & Jasper, J. D. (2000). Information processing at successive stages of decision making: Need for cognition and inclusion–exclusion effects. *Organizational Behavior and Human Decision Processes*, *82*(2), 171–193.

Li, Y. & Epley, N. (2009). When the best appears to be saved for last: Serial position effects on choice. *Journal of Behavioral Decision Making*, *22*(4), 378–389.

Liddell, T. M. & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, *79*, 328–348.

Lidén, M., Gräns, M., & Juslin, P. (2018). The presumption of guilt in suspect interrogations: Apprehension as a trigger of confirmation bias and debiasing techniques. *Law and Human Behavior*, *42*(4), 336–354.

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.

Liu, J., Thomas, J. M., & Higgs, S. (2019). The relationship between social identity, descriptive social norms and eating intentions and behaviors. *Journal of Experimental Social Psychology*, *82*, 217–230.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109.

Lynn, M. (1989). Scarcity effects on desirability: Mediated by assumed expensiveness? *Journal of Economic Psychology*, *10*(2), 257–274.

Lynn, M. (1992). Scarcity's enhancement of desirability: The role of naive economic theories. *Basic and Applied Social Psychology*, *13*(1), 67–78.

Lynn, M. & Harris, J. (1997). Individual differences in the pursuit of self-uniqueness through consumption. *Journal of Applied Social Psychology*, *27*(21), 1861–1883.

McFadden, D., Machina, M. J., & Baron, J. (1999). Rationality for economists? In *Elicitation of preferences* (pp. 73–110). Springer.

Melnyk, V., Herpen, E. v., Fischer, A. R., & van Trijp, H. (2011). To think or not to think: The effect of cognitive deliberation on the influence of injunctive versus descriptive social norms. *Psychology & Marketing*, *28*(7), 709–729.

Melnyk, V., Van Herpen, E., Jak, S., & Van Trijp, H. C. (2019). The mechanisms of social norms' influence on consumer decision making. *Zeitschrift für Psychologie*, *227*(1), 4–17.

Miyapuram, K. P., Tobler, P. N., Gregorios-Pippas, L., & Schultz, W. (2012). BOLD responses in reward regions to hypothetical and imaginary monetary rewards. *NeuroImage*, *59*(2), 1692–1699.

Morewedge, C. K. & Giblin, C. E. (2015). Explanations of the endowment effect: An integrative review. *Trends in Cognitive Sciences*, *19*(6), 339–348.

Morewedge, C. K., Shu, L. L., Gilbert, D. T., & Wilson, T. D. (2009). Bad riddance or good rubbish? Ownership and not loss aversion causes the endowment effect. *Journal of Experimental Social Psychology*, *45*(4), 947–951.

Mukherjee, S. & Srinivasan, N. (2013). Attention in preferential choice. In *Progress in brain research* (Vol. 202, pp. 117–134). Amsterdam: Elsevier.

Nakhaeizadeh, S., Dror, I. E., & Morgan, R. M. (2014). Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. *Science & Justice*, *54*(3), 208–214.

Nayakankuppam, D. & Mishra, H. (2005). The endowment effect: Rose-tinted and dark-tinted glasses. *Journal of Consumer Research*, *32*(3), 390–395.

Neighbors, C., LaBrie, J. W., Hummer, J. F., Lewis, M. A., Lee, C. M., Desai, S., ... Larimer, M. E. (2010). Group identification as a moderator of the relationship between perceived social norms and alcohol consumption. *Psychology of Addictive Behaviors*, *24*(3), 522–528.

Newell, B. R. & Bröder, A. (2008). Cognitive processes, models and metaphors in decision research. *Judgment and Decision Making*, *3*(3), 195–204.

Newell, B. R. & Shanks, D. R. (2007). Recognising what you like: Examining the relation between the mere-exposure effect and recognition. *European Journal of Cognitive Psychology*, *19*(1), 103–118.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220.

Nicolle, A., Fleming, S. M., Bach, D. R., Driver, J., & Dolan, R. J. (2011). A regret-induced status quo bias. *The Journal of Neuroscience*, *31*(9), 3320–3327.

Nissani, M. & Hoefler-Nissani, D. M. (1992). Experimental studies of belief dependence of observations and of resistance to conceptual change. *Cognition and Instruction*, *9*(2), 97–111.

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.

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and Social Psychology Bulletin*, *34*(7), 913–923.

Novemsky, N. & Kahneman, D. (2005). The boundaries of loss aversion. *Journal of Marketing Research*, *42*(2), 119–128.

Ordóñez, L. D., Connolly, T., & Coughlan, R. (2000). Multiple reference points in satisfaction and fairness assessment. *Journal of Behavioral Decision Making*, *13*(3), 329–344.

Oruc, R. (2015). *The Effects of Product Scarcity on Consumer Behavior: A Meta-Analysis* (Doctoral dissertation, Europa-Universität Viadrina Frankfurt).

Ounjai, K., Kobayashi, S., Takahashi, M., Matsuda, T., & Lauwereyns, J. (2018, November 15). Active confirmation bias in the evaluative processing of food images. *Scientific Reports*, *8*(1), 16864.

Pachur, T. & Scheibehenne, B. (2012). Constructing preference from experience: The endowment effect reflected in external information search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1108–1116.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgement and Decision Making*, *5*(5), 411–419.

Parducci, A. (1974). Contextual effects: A range-frequency analysis. *Handbook of perception*, *2*, 127–141.

Park, D.-H., Lee, J., & Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. *International Journal of Electronic Commerce*, *11*(4), 125–148.

Passafaro, P., Livi, S., & Kosic, A. (2019). Local norms and the theory of planned behaviour: Understanding the effects of spatial proximity on recycling intentions and self-reported ecological behaviour. *Frontiers in Psychology*, *10*, 1–11.

Pettibone, J. C. & Wedell, D. H. (2000). Examining models of nondominated decoy effects across judgment and choice. *Organizational Behavior and Human Decision Processes*, *81*(2), 300–328.

Pettigrew, T. F. & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783.

Platow, M. J., Haslam, S. A., Both, A., Chew, I., Cuddon, M., Goharpey, N., … Grace, D. M. (2005). "It's not funny if they're laughing": Self-categorization, social influence, and responses to canned laughter. *Journal of Experimental Social Psychology*, *41*(5), 542–550.

Plott, C. R. & Zeiler, K. (2007). Exchange asymmetries incorrectly interpreted as evidence of endowment effect theory and prospect theory? *The American Economic Review*, *97*(4), 1449–1466.

Postmes, T., Haslam, S. A., & Jans, L. (2013). A single-item measure of social identification: Reliability, validity, and utility. *British Journal of Social Psychology*, *52*(4), 597–617.

Pratkanis, A. R. & Farquhar, P. H. (1992). A brief history of research on phantom alternatives: Evidence for seven empirical generalizations about phantoms. *Basic and Applied Social Psychology*, *13*(1), 103–122.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, *32*, 122–136.

Pryor, C. G., Perfors, A., & Howe, P. D. L. (2018). Reversing the endowment effect. *Judgement and Decision Making*, *13*(3), 275–286.

Pryor, C. G., Perfors, A., & Howe, P. D. L. (2019a). Conformity to the descriptive norms of people with opposing political or social beliefs. *PLoS ONE*, *14*(7), e0219464.

Pryor, C. G., Perfors, A., & Howe, P. D. L. (2019b). Even arbitrary norms influence moral decision-making. *Nature Human Behaviour*, *3*, 57–62.

R Core Team. (2016). *R: A language and environment for statitical computing of Work*. Vienna, Austria: R foundation for statistical computing.

Rajsic, J., Wilson, D. E., & Pratt, J. (2015). Confirmation bias in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(5), 1353–1364.

Ranzini, M., Dehaene, S., Piazza, M., & Hubbard, E. M. (2009). Neural mechanisms of attentional shifts due to irrelevant spatial and numerical cues. *Neuropsychologia*, *47*(12), 2615–2624.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281.

Reynolds, K. J. (2019). Social norms and how they impact behaviour. *Nature Human Behaviour*, *3*(1), 14–15.

Rimal, R. N. (2008). Modeling the relationship between descriptive norms and behaviors: A test and extension of the theory of normative social behavior (TNSB). *Health Communication*, *23*(2), 103–116.

Rimal, R. N., Lapinski, M. K., Cook, R. J., & Real, K. (2005). Moving toward a theory of normative influences: How perceived benefits and similarity moderate the impact of descriptive norms on behaviors. *Journal of Health Communication*, *10*(5), 433–450.

Rimal, R. N. & Real, K. (2003). Understanding the influence of perceived norms on behaviors. *Communication Theory*, *13*(2), 184–203.

Rimal, R. N. & Real, K. (2005). How behaviors are influenced by perceived norms: A test of the theory of normative social behavior. *Communication Research*, *32*(3), 389–414.

Ritov, I. & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making*, *3*(4), 263–277.

Ritov, I. & Baron, J. (1992). Status-quo and omission biases. *Journal of Risk and Uncertainty*, *5*(1), 49–61.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological Review*, *108*(2), 370–392.

Roos, P., Gelfand, M., Nau, D., & Lun, J. (2015). Societal threat and cultural variation in the strength of social norms: An evolutionary basis. *Organizational Behavior and Human Decision Processes*, *129*, 14–23.

Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes*, *66*(1), 102–110.

Russo, J. E., Meloy, M. G., & Medvec, V. H. (1998). Predecisional Distortion of Product Information. *Journal of Marketing Research*, *35*(4), 438–452.

Samuelson, W. & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty*, *1*(1), 7–59.

Sarnoff, I. & Zimbardo, P. G. (1961). Anxiety, fear, and social isolation. *The Journal of Abnormal and Social Psychology*, *62*(2), 356–363.

Schmidt, U. & Traub, S. (2002). An Experimental Test of Loss Aversion. *Journal of Risk and Uncertainty*, *25*(3), 233–249.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, *18*(5), 429–434.

Schweitzer, M. (1994). Disentangling status quo and omission effects: An experimental analysis. *Organizational Behavior and Human Decision Processes*, *58*(3), 457–476.

Sehnert, S., Franks, B., Yap, A. J., & Higgins, E. T. (2014). Scarcity, engagement, and value. *Motivation and Emotion*, *38*(6), 823–831.

Sen, S. (1998). Knowledge, information mode, and the attraction effect. *Journal of Consumer Research*, *25*(1), 64–77.

Senecal, S. & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, *80*(2), 159–169.

Shavitt, S. (1990). The role of attitude objects in attitude functions. *Journal of Experimental Social Psychology*, *26*(2), 124–148.

Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology (Columbia University)*.

Sherif, M. (1936). *The psychology of social norms*. New York & London: Harper & Brothers Publishing.

Shevchenko, Y., Von Helversen, B., & Scheibehenne, B. (2014). Change and status quo in decisions with defaults: The effect of incidental emotions depends on the type of default. *Judgment and Decision making*, *9*(3), 287–296.

Shi, L. & Wang, K. (2008). An Laboratory Experiment for Comparing Effectiveness of Three Types of Online Recommendations. *Tsinghua Science & Technology*, *13*(3), 293–299.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317–1322.

Shu, S. B. & Peck, J. (2011). Psychological ownership and affective reaction: Emotional attachment process variables and the endowment effect. *Journal of Consumer Psychology*, *21*(4), 439–452.

Simion, C. & Shimojo, S. (2007). Interrupting the cascade: Orienting contributes to decision making even in the absence of visual stimulation. *Perception & Psychophysics*, *69*(4), 591–595.

Simon, D., Krawczyk, D. C., & Holyoak, K. J. (2004). Construction of preferences by constraint satisfaction. *Psychological Science*, *15*(5), 331–336.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*(2), 158–174.

Simonson, I. & Kivetz, R. (2018). Bringing (Contingent) Loss Aversion Down to Earth—A Comment on Gal & Rucker's Rejection of "Losses Loom Larger Than Gains". *Journal of Consumer Psychology*, *28*(3), 517–522.

Slovic, P. (1995). The construction of preference. *American Psychologist*, *50*(5), 364–371.

Smerecnik, C. M., Mesters, I., Kessels, L. T., Ruiter, R. A., De Vries, N. K., & De Vries, H. (2010). Understanding the positive effects of graphical risk information on comprehension: Measuring attention directed to written, tabular, and graphical risk information. *Risk Analysis*, *30*(9), 1387–1398.

Smith, J. R. & Terry, D. J. (2003). Attitude-behaviour consistency: The role of group norms, attitude accessibility, and mode of behavioural decision-making. *European Journal of Social Psychology*, *33*(5), 591–608.

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, *44*(3), 408–463.

Snyder, C. R. & Fromkin, H. L. (2012). *Uniqueness: The human pursuit of difference*. New York: Plennum Press.

Sokol-Hessner, P., Camerer, C. F., & Phelps, E. A. (2012). Emotion regulation reduces loss aversion and decreases amygdala responses to losses. *Social Cognitive and Affective Neuroscience*, *8*(3), 341–350.

Stafford, T. & Grimes, A. (2012). Memory Enhances the Mere Exposure Effect. *Psychology & Marketing*, *29*(12), 995–1003.

Steffel, M. & Williams, E. F. (2018). Delegating decisions: Recruiting others to make choices we might regret. *Journal of Consumer Research*, *44*(5), 1015–1032.

Stok, F. M., De Ridder, D. T. D., De Vet, E., & De Wit, J. B. F. (2014). Don't tell me what I should do, but what others do: The influence of descriptive and injunctive peer norms on fruit consumption in adolescents. *British Journal of Health Psychology*, *19*(1), 52–64.

Sütterlin, B., Brunner, T. A., & Opwis, K. (2008). Eye-tracking the cancellation and focus model for preference judgments. *Journal of Experimental Social Psychology*, *44*(3), 904–911.

Symons, C. S. & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*(3), 371–394.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*(5), 96–103.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178.

Talluri, B. C., Urai, A. E., Tsetsos, K., Usher, M., & Donner, T. H. (2018). Confirmation bias through selective overweighting of choice-consistent evidence. *Current Biology*, *28*(19), 3128–3135.

Tannenbaum, D., Fox, C. R., & Rogers, T. (2017). On the misplaced politics of behavioural policy interventions. *Nature Human Behaviour*, *1*(7), 1–7.

Terry, D. J. & Hogg, M. A. (1996). Group norms and the attitude-behavior relationship: A role for group identification. *Personality and Social Psychology Bulletin*, *22*(8), 776–793.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, *1*(1), 39–60.

Tom, G., Nelson, C., Srzentic, T., & King, R. (2007). Mere exposure and the endowment effect on consumer decision making. *The Journal of Psychology*, *141*(2), 117–125.

Trueblood, J. S. (2015). Reference point effects in riskless choice without loss aversion. *Decision*, *2*(1), 13–26.

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014a). 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. (2014b). The multi-attribute linear ballistic accumulator model of decision-making. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*, 3581–3586.

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.

Trueblood, J. S. & Pettibone, J. C. (2015). The phantom decoy effect in perceptual decision making. *Journal of Behavioral Decision Making*, *30*(2), 157–167.

Tsetsos, K., Gao, J., McClelland, J. L., & Usher, M. (2012). Using time-varying evidence to test models of decision dynamics: Bounded diffusion vs. the leaky competing accumulator model. *Frontiers in Neuroscience*, *6*, 1–17.

Turner, J. C. (2005). Explaining the nature of power: A three-process theory. *European Journal of Social Psychology*, *35*(1), 1–22.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, UK: Basil Blackwell.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Tversky, A. & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061.

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.

Usher, M. & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*(3), 757–769.

Vinnell, L. J., Milfont, T. L., & McClure, J. (2019). Do social norms affect support for earthquake-strengthening legislation? Comparing the effects of descriptive and injunctive norms. *Environment and Behavior*, *51*(4), 376–400.

von Neumann, J. & Morgenstern, O. (1953). *Theory of games and economic behaviour* (Third). Princeton: Princeton University Press.

Voyer, B. (2015). 'Nudging' behaviours in healthcare: Insights from behavioural economics. *British Journal of Healthcare Management*, *21*(3), 130–135.

Wang, M. & Fischbeck, P. S. (2004). Incorporating framing into prospect theory modeling: A mixture-model approach. *Journal of Risk and Uncertainty*, *29*(2), 181–197.

Wellen, J. M., Hogg, M. A., & Terry, D. J. (1998). Group norms and attitude–behavior consistency: The role of group salience and mood. *Group Dynamics: Theory, Research, and Practice*, *2*(1), 48–56.

Wenzel, M. (2005). Misperceptions of social norms about tax compliance: From theory to intervention. *Journal of Economic Psychology*, *26*(6), 862–883.

Wenzel, M. (2019). *Misperceptions of social norms about tax compliance (2): A field experiment* (No. 7). Australian National University. Canberra.

White, K. & Dahl, D. W. (2006). To Be or Not Be? The Influence of Dissociative Reference Groups on Consumer Preferences. *Journal of Consumer Psychology*, *16*(4), 404–414. doi:https://doi.org/10.1207/s15327663jcp1604_11

Willemsen, M. C., Böckenholt, U., & Johnson, E. J. (2011). Choice by value encoding and value construction: Processes of loss aversion. *Journal of Experimental Psychology: General*, *140*(3), 303–324.

Williams, C. C., Saffer, B. Y., McCulloch, R. B., & Krigolson, O. E. (2016). The scarcity heuristic impacts reward processing within the medial-frontal cortex. *NeuroReport*, *27*(7), 522–526.

Wingenfeld, K., Mensebach, C., Driessen, M., Bullig, R., Hartje, W., & Beblo, T. (2006). Attention Bias towards Personally Relevant Stimuli: The Individual Emotional Stroop Task. *Psychological Reports*, *99*(3), 781–793.

Winkielman, P. & Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology*, *81*(6), 989–1000.

Winkielman, P., Schwarz, N., Fazendeiro, T., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In *The Psychology of Evaluation: Affective Processes in Cognition and Emotion* (pp. 189–217). New Jersey: Lawrence Erlbaum Associates.

Wu, W.-Y., Lu, H.-Y., Wu, Y.-Y., & Fu, C.-S. (2012). The effects of product scarcity and consumers' need for uniqueness on purchase intention. *International Journal of Consumer Studies*, *36*(3), 263–274.

Xiang Fang, Surendra Singh, & Rohini Ahluwalia. (2007). An examination of different explanations for the mere exposure effect. *Journal of Consumer Research*, *34*(1), 97–103.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2), 1–27.

Zhu, M. & Ratner, R. K. (2015). Scarcity polarizes preferences: The impact on choice among multiple items in a product class. *Journal of Marketing Research*, *52*(1), 13–26.

Appendix A

Distinguishing theories based on undesirable choice stimuli

In this appendix we briefly review a wide range of different theories that have been applied to explain various choice biases. We outline how each of these theories explain the relevant choice biases to which they have been applied and whether they predict these choice biases to reverse or remain consistent when choosing between undesirable, rather than desirable, choice stimuli.

A.1 Salience theory

Summary A target item (e.g. an endowed item) is believed to be evaluated first and compared to a decision maker's previous state (Bordalo et al., 2012a). Assuming the previous state is neutral, the target item will stand out, amplifying its perceived value. When subsequently evaluating similarly-desirable alternatives, the decision maker evaluates them in the context of the target item, such that these alternatives do not stand out as much as the target item and thus, have a more neutral evaluation.

Prediction **Reversal for undesirable stimuli**. Due to the value of the target item being amplified relative to alternatives, a desirable target item will be preferred due to its desirability being amplified. Meanwhile, an undesirable target item will have its undesirability amplified, such that it is preferred less.

A.2 Associative accumulation model

Summary Proposes that we weight attributes more when they have high activation (Bhatia, 2013). Activation is due to the available choice options (and especially salient choice options) scoring highly on that attribute. Due to target items generally being more salient (or having a decoy that scores highly on a similar attribute in the case of the attraction, phantom decoy and compromise effects), attributes the target item scores highly on are weighted more.

Prediction **Reversal for undesirable stimuli**. If the target item scores highly on desirable attributes, then increased weighting of those attributes will boost preference for the target item. In contrast, if the target item scores highly on undesirable attribute, increased weighting of these attributes will make the target item appear worse, leading to reversals of choice biases.

A.3 Query theory

Summary Proposes that we tend to evaluate the reasons to keep an endowed option before coming up with reasons to not keep an endowed option (Johnson et al., 2007). The initial process of evaluating reasons to keep an endowed option is thought to inhibit subsequent thoughts about reasons to not keep that option, leading to the endowment effect.

Prediction **No reversal for undesirable stimuli**. Query theory assumes that people evaluate more reasons to choose an endowed option relative to an alternative, independent of the absolute value of these options. Thus, the endowed option is assumed to be favoured regardless of desirability.

A.4 Psychological ownership theory

Summary Proposes that the endowment effect occurs because ownership of an item inherently boosts perception of that item, either because it becomes intertwined with the individual's self-identity (Gawronski, Bodenhausen, & Becker, 2007) or the individual becomes emotionally attached to the item (Shu & Peck, 2011)

Prediction **No reversal for undesirable stimuli**. Provided we assume that individuals become similarly attached to undesirable goods as they do to desirable goods, this attachment is predicted to increase preference for the owned item regardless of desirability.

A.5 Self-referential memory effects

Summary It has been found that people have better memory for information when it can be linked to themselves (Symons & Johnson, 1997). This extends to people having better memory for items they own (Cunningham, Turk, Macdonald, & Macrae, 2008). Having stronger memory for an item when making a decision may increase its perceived valence such that a desirable owned item is preferred because its desirability is better remembered when making a decision (Morewedge & Giblin, 2015).

Prediction **Reversal for undesirable stimuli**. Stronger memory for an owned item will increase its preference share when it is desirable due to stronger memory of its desirable features. On the other hand, an undesirable item will have undesirable features and thus, stronger memory of these features will decrease the owned items preference share.

A.6 Prospect Theory

Summary Proposes that we evaluate options relative to a reference point (Kahneman & Tversky, 1979). Manipulations such as endowing an option are believed to establish that option as a reference point. Losses relative to this reference point loom larger than relative gains, such that decision makers tend to prefer options that are most similar to the reference point.

Prediction **No reversal for undesirable stimuli**. Prospect theory assumes that the gains and losses of an option relative to a reference point determine its perceived value. Thus, regardless of whether the options are both desirable or both undesirable, if the relative value of the options to the reference point remain consistent, the decision maker should continue to be biased towards favouring options that are similar to the reference point.

A.7 Possession loss aversion

Summary Proposes that the subjective experience of losing possession of an owned item has a stronger valence than the experience of gaining possession of an item (Brenner et al., 2007).

Prediction **Reversal for undesirable stimuli**. Losing a desirable endowed item is believed to be a particularly negative experience, more so than gaining a new desirable item is positive. This predicts that people will want to keep possession of desirable endowments. In contrast, if the endowment is undesirable then losing possession of it is particularly positive, outweighting the potential negativity of gaining possession of an alternative undesirable item. Thus, an endowment reversal is predicted for undesirable choice options.

A.8 Range-frequency theory

Summary Explains the attraction effect based on the idea that an inferior decoy either extends the range of values on the competitor favoured attribute and/or increases the frequency of values on the target preferred attribute (Huber et al., 1982; Parducci, 1974). Extending the range of the competitor preferred attribute is believed to decrease the apparent disadvantages of the target on that attribute, thereby boosting its preference share. Increasing the frequency of values on the target preferred attribute is expected to exaggerate the advantages of the target item, thereby boosting its preference.

Prediction **No reversal for undesirable stimuli**. Regardless of whether the options are desirable or undesirable, the inferior decoy will still extend the range of the competitor-preferred attribute or increase the frequency of values on the target-preferred attribute, meaning that the absolute desirability of the options is irrelevant.

A.9 Leaky Competing Accumulator model

Summary Proposes that options are compared in a pairwise manner, with a random option acting as the reference point during each comparison (Usher & McClelland, 2001; Usher & McClelland, 2004). During each of these pairwise comparisons, an accumulator shifts in favour of or against the non-reference option based on its value relative to the reference option, where losses are weighted more than gains. This explains the attraction, phantom decoy and compromise effect based on the fact that the decoy option is more similar to the target option than the competitor option. Because losses are weighted more than gains, the larger relative losses of the competitor to the decoy than the target to the decoy result in the target being preferred over the competitor.

Prediction **No reversal for undesirable stimuli**. Only focuses on the relative value of options to each other, rather than their absolute value, such that whether the items are desirable or undesirable is considered irrelevant.

A.10 Multialternative Decision Field Theory

Summary Proposes that at any point in time, all choice options are compared on a sampled attribute, where if the valence of an item is greater than the average valence of other items on that attribute, an accumulator shifts in favour of that item (Roe et al., 2001). This is said to be affected by lateral inhibition, where similar items can interact. An item with stronger preference (as represented by the accumulator having shifted more strongly in favour of that option) inhibits preference for similar but lower preference items. Meanwhile, lateral inhibition also means that lower preference items can boost preference for the similar but superior item. This explains the attraction effect due to the low preference that develops for the dominated decoy boosting preference for the similar but superior target item. The similarity effect is explained due to similar items having positively correlated preference such that they compete for preference share. The compromise effect is attributed to lateral inhibition causing the compromise option to become negatively correlated with both of the extreme alternatives, boosting its preference share.

Prediction **No reversal for undesirable stimuli**. Shifts in the decision accumulators are based on the value of options relative to the average value of other options on a sample attribute. This comparison to average scores means that the absolute value of the options is partialled out and becomes inconsequential to predicted preferences.

A.11 Multi-attribute Linear Ballistic Accumulator model

Summary Consists of two stages, a front-end item comparison process and a back-end preference accumulation process (Trueblood, Brown, & Heathcote, 2014b). The back-end process simply involves linear ballistic accumulators (representing each choice option) racing towards a threshold at a drift rate determined by the front-end process, which distinguishes the subjective value of the options. The front-end process consists of comparing each item to all other items. The subjective value of options based on each comparison is determined by mapping objective values to psychological magnitudes. It is proposed that these values are mapped to a curve which typically favours more central attribute values. The subjective values from these comparisons are then weighted based on how likely these comparisons are to be made. It is proposed that comparisons are more likely for items that are more similar (and thus, more difficult to distinguish). Depending on how parameters are set, this model can explain the attraction, compromise and similarity decoy effects. For example, the attraction effect is attributed to comparisons between similar items being more common. This leads to favourable comparisons for the target option to the similar but inferior decoy being common and boosting preference for the target option.

Prediction **No reversal for undesirable stimuli**. The model assumes that all comparisons are based on the relative value of options to each other. Thus, the model does not distinguish between options that are desirable or undesirable, provided their relative value remains constant.

A.12 Justification theory

Summary Proposes that people want to be able to justify their decision to themselves or others (Simonson, 1989). Various manipulations can make an option seem more justifiable than others. For example, in the attraction effect, domination of the target over the decoy makes it easier to justify choosing the target option. For the similarity effect, the similarity of the decoy to the competitor option makes it hard to justify choosing one option over the other, such that the target option is instead relatively easier to justify choosing.

Prediction **No reversal for undesirable stimuli**. Though a fairly broad idea, the way it has been applied to explaining effects generally indicates that no reversal is predicted for undesirable choice situations compared to desirable choice situations. For example, even if the options are undesirable, a target item still has a dominance relationship over a decoy, such that it is easier to justify choosing it, leading to the attraction effect.

A.13 Similarity substitution

Summary Proposes that people want to choose the best item. Failing that, they will choose whichever available option is most similar to the best item (Pettibone & Wedell, 2000). Thus, the phantom decoy effect is attribute to people wanting to choose the superior phantom decoy option but settling for the inferior but similar target option.

Prediction **No reversal for undesirable stimuli**. Regardless of whether the options are desirable or undesirable, the phantom decoy option will still be the best of the available options. Thus, similarity substitution predicts that people will settle for the similar but inferior target option regardless of absolute desirability.

A.14 Commodity theory

Summary Explains the scarcity bias based on the idea that scarcity of a product makes it seem more valuable (Lynn, 1989) or unique (Snyder & Fromkin, 2012).

Prediction **No reversal for undesirable stimuli**. Proposes that scarcity provides a direct boost to the perceived desirability of an option, regardless of whether that option is, overall, desirable or undesirable.

A.15 Social sanction account

Summary Proposes that we conform to social norms (e.g. the injunctive or descriptive norms) in an effort to elicit positive responses (positive social sanctions) from others and avoid the potential for people to respond negatively (negative social sanctions) for deviating from the social norm (Ajzen & Fishbein, 1980; Bendor & Swistak, 2001).

Prediction **No reversal for undesirable stimuli**. The threat of social sanctions is independent of the desirability of the choice options. Thus, a decision maker should be similarly affected by the potential for social sanctions regardless of whethe the choice options are desirable or undesirable.

A.16 Informational account

Summary Proposes that individuals conform to descriptive norms because the choices of others are indicative of "effective and adaptive action" (Cialdini et al., 1990, p. 1015). If most other people favour a certain option it is likely because they deemed that option to be best and thus, that option is also likely to be best for the decision maker.

Prediction **No reversal for undesirable stimuli**. Regardless of whether the options are desirable or undesirable, if most other people prefer a certain action it should be similarly indicative of which option is best. Thus, descriptive norms offer the same information and are expected to be followed regardless of desirability.

A.17 Self-categorization theory

Summary Proposes that norm effects occur when an individual's identity as a member of a particular social group is particularly salient (Turner et al., 1987; Terry & Hogg, 1996). When this is the case, the individual is believed to want to minimise ingroup differences and maximise intergroup differences. Thus, to minimise ingroup differences, individuals conform to the social norms of their ingroup.

Prediction **No reversal for undesirable stimuli**. Even if a behaviour is undesirable, it can still be seen as an aspect of group identity. Thus, regardless of desirability, self-categorization theory predicts that people will be more likely to favour an option when it is favoured by their ingroup.

A.18 Perceptual fluency misattribution theory

Summary Proposes that the mere exposure effect occurs because prior exposure to a stimulus makes it easier to process (Jacoby, Kelley, & Dywan, 1989; Bornstein & D'Agostino, 1994). This easier processing (fluency) is thought to be mistakenly attributed to liking the stimulus.

Prediction **No reversal for undesirable stimuli**. Regardless of whether the stimulus is desirable or undesirable, exposure to it should make it easier to process and thus, misattribution of this fluency to greater liking should occur even for undesirable stimuli.

A.19 Uncertainty reduction

Summary When evaluating a stimulus, uncertainty and unfamiliarity with the stimulus are negative experiences (Bornstein, 1989). This theory proposes that the less novel a stimulus is, the less of this negative uncertainty occurs. Thus, the mere exposure effect is attributed to the idea that if an item has been seen before, it will be less novel, reducing the experience of uncertainty and thus, making it appear more desirable (i.e. less aversive).

Prediction **No reversal for undesirable stimuli**. Exposing a stimulus should reduce uncertainty in a consistent manner regardless of whether the stimuli are desirable or undesirable.

A.20 Hedonic fluency

Summary Explains the mere exposure effect based on the idea that exposure leads to greater perceptual fluency (easier processing). Easier processing leads to easier recognition (Winkielman et al., 2003). Recognition is then believed to increase positive affect towards an item for various reasons such as recognition indicating that a stimulus is harmless or recognition being inherently rewarding.

Prediction **No reversal for undesirable stimuli**. Proposes that the experience of recognition is inherently rewarding and accordingly this recognition directly increases liking of an exposed item, regardless of its baseline desirability.

A.21 Cancellation and focus theory

Summary Explains the serial order effect based on the idea that features are given more attention when they are salient (Houston et al., 1989; Houston & Sherman, 1995). When evaluating the first item in a serial order effect paradigm, the decision maker typically has no past context and thus all of the features are similarly salient. In contrast, when evaluating the second item, any shared features with the first item are familiar and stand out less. Instead, unique features become salient such that they are given greater weight in the decision. When the options differ on unique positive features, then having these positive features of the latter item stand out increases its preference share. Meanwhile, if the options differ on negative features then having these unique negative features of the latter item stand out decreases its preference share.

Prediction **Reversal for undesirable stimuli**. Because the features of the last presented option stand out, cancellation and focus theory predicts that people will favour the first presented option when the options differ on undesirable features but the last presented option when the options differ on desirable features.

A.22 Memory effects

Summary Proposes that serial order effects occur because memory of earlier presented items fades such that the earlier presented options have a lower valence than the later presented options when making a decision (Houston et al., 1989; Li & Epley, 2009).

Prediction **Reversal for undesirable stimuli**. If the options are desirable, then the valence of the earlier option fading will mean it is perceived as less desirable than the more recent, higher perceived valence option. In contrast, if the options are undesirable, then having a weaker memory of how undesirable the earlier option was will increase its preference share.

A.23 Pre-decisional information distortion

Summary Proposes that we process available information such that it favours an early leader (Russo et al., 1998; Lord et al., 1979). For example, in the leader driven primacy effect, establishing one option as an early leader is thought to bias subsequent processing of information such that it too more strongly favours the leader, even if that information is actually neutral (Carlson et al., 2006).

Prediction **No reversal for undesirable stimuli**. Even if the choice options are undesirable, pre-decisional information distortion predicts that one option being favoured early on over another option will bias information processing in favour of that early leader.

Appendix B

Alternative value function for Chapter 4

In order to assess the extent to which the predictions of the model outlined in this paper were dependent on the specific value function used, we refit the model using the value function employed by Usher and McClelland, 2004 in their Leaky Competing Accumulator (LCA) model, as shown in Equation (B.1). As can be seen in Figure B.1, the predictions of the model with the LCA value function were essentially equivalent to the predictions of the model presented in this paper. The parameter estimates obtained in this case were $\gamma = 0.637$, $\sigma = 0.792$ and $\delta = 0.172$.

$$\mu(z) = \begin{cases} log(1+z), & \text{if } z \geq 0 \\ -\{(log(1+|z|) + [log(1+|z|)]^2\}, & \text{if} z < 0 \end{cases} \tag{B.1}$$

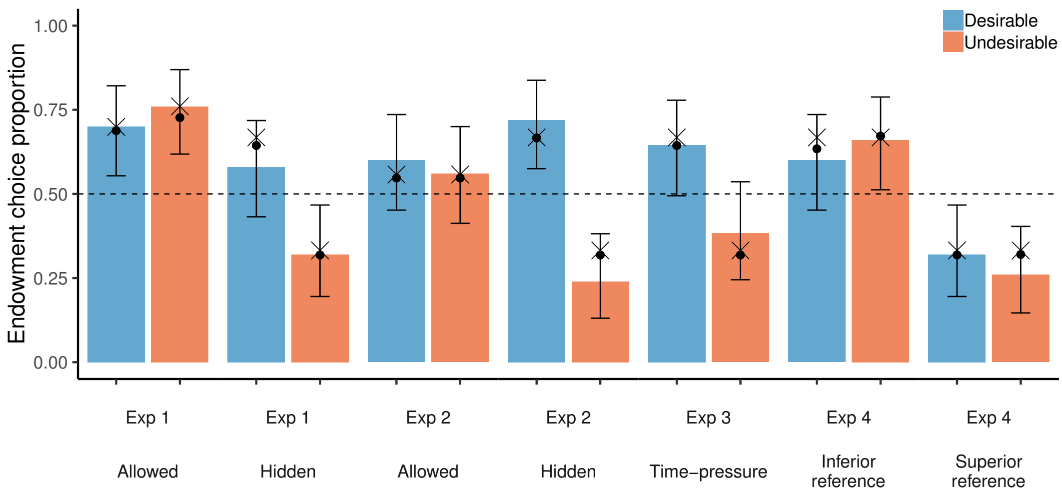

FIGURE B.1: Observed endowment preference share (error bars represent 95%CIs) for Experiments 1A, 1B, 2 and 3. The crosses represent the predicted endowment preference share of the model using the value function from Equation (4.4) on page 54. The dots represent the predicted endowment preference share of the model using the LCA value function from Equation (B.1).

Appendix C

Alternative approach to modelling information integration for Chapter 4

In this paper we presented a model that set the current reference point to either the current endowed option or the previous reference state with a certain probability. We present here an alternative method of combining the previous reference state and current endowed state to assess whether the predictions of our model are consistent across assumptions about how information is integrated.

Here we treat both the previous reference state and the current endowed state as probability distributions of different outcomes. We then set the current reference state equal to a weighted average of the probabilities in each of these probability distributions. Thus, the current reference point, r, is equal to a probability distribution across all of the possible outcomes in both the current endowed state, ε, and the previous reference state, s, where the probability of each possible outcome is simply equal to the weighted average of the probability of that outcome in the previous reference state and the current reference state. This is shown in Equation (C.1), where p_i^x represents the probability of outcome i in probability distribution x and γ represents the impact that the current endowment has on expectations relative to the impact that the previous referent point has, with the restriction that $0 < \gamma < 1$. The parameter estimates obtained when this model was fit to our data were $\gamma = 0.899$, $\sigma = 1.921$ and $\delta = 0.198$.

$$p_i^r = \gamma p_i^\varepsilon + (1 - \gamma) p_i^s \tag{C.1}$$

As an example, assume the current endowment is a 50% chance of winning \$0.20 and a 50% chance of winning nothing, whilst the previous reference state was a 25% chance of winning \$0.10 and a 75% chance of winning nothing. If $\gamma = 0.5$, then the current reference state would be a 25% chance of winning \$0.20, a 12.5% chance of winning \$0.10 and a 62.5% chance of winning nothing. As can be seen from Figure C.1 on the following page, this model fits our data approximately as well as the previous model, which provides further evidence that our results can be explained in terms of loss aversion across a number of reasonable model assumptions.

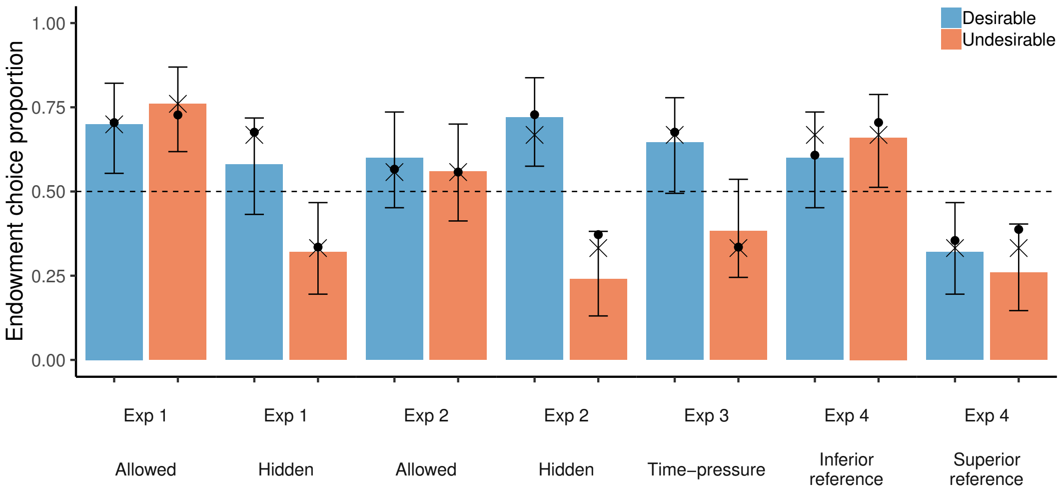

FIGURE C.1: Observed endowment preference share (error bars represent 95% CIs) across Experiments 1A, 1B, 2 and 3. The crosses represent the predicted endowment preference share of the model described in this paper. The dots represent the predicted endowment preference share of the model presented in this appendix, which combines the previous reference state and the current endowed state according to Equation (C.1) on the preceding page.

Appendix D

Issue Statements From Chapter 8

Following is the list of issues and corresponding statements that participants were presented with in Experiments 1 and 3 of Chapter 8. Figure D.1 on the next page and Figure D.2 on the following page show the joint frequency with which participants chose each issue as the one they cared about most, and the extent to which they then agreed or disagreed with the subsequent statement about that issue.

- **Gun control**: "Adults should have the right to carry a concealed handgun"
- **Feminism**: "Feminism is important and beneficial to modern society"
- **Donald Trump**: "Donald Trump being president is good for the United States at this time"
- **Immigration and Dreamers**: "Dreamers (undocumented immigrants who came to the US as children) should be allowed to stay in the United States"
- **Transgender rights**: "Transgender people should be allowed to use the bathrooms of the gender they identify as"
- **Drug legalization**: "Possession of drugs should be legalized"
- **Colin Kaepernick kneeling during the national anthem**: "Colin Kaepernick was wrong to kneel during the national anthem"
- **Buying and wearing fur**: "Buying and wearing fur is wrong"
- **Taxing religious organization**: "Religious organizations should be taxed"

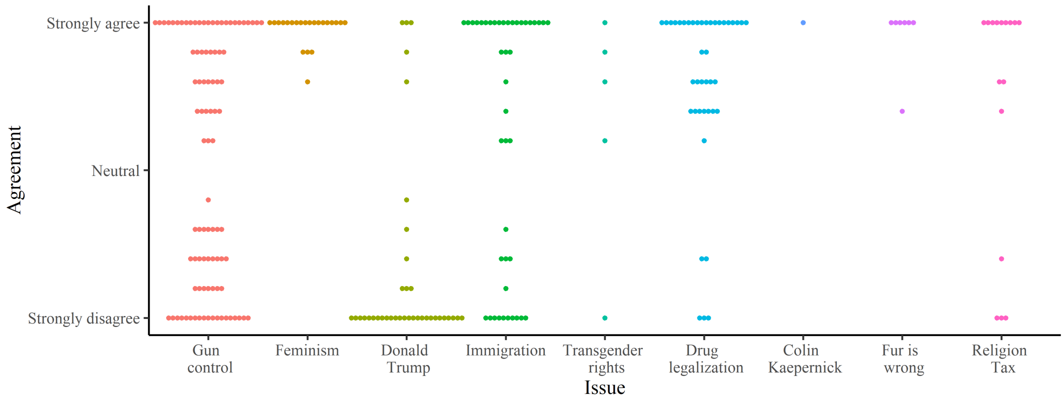

FIGURE D.1: Dot plot of the joint frequency of participants caring most about an issue and the extent to which they then agreed with the statement about that issue in Experiment 1 of Chapter 8.

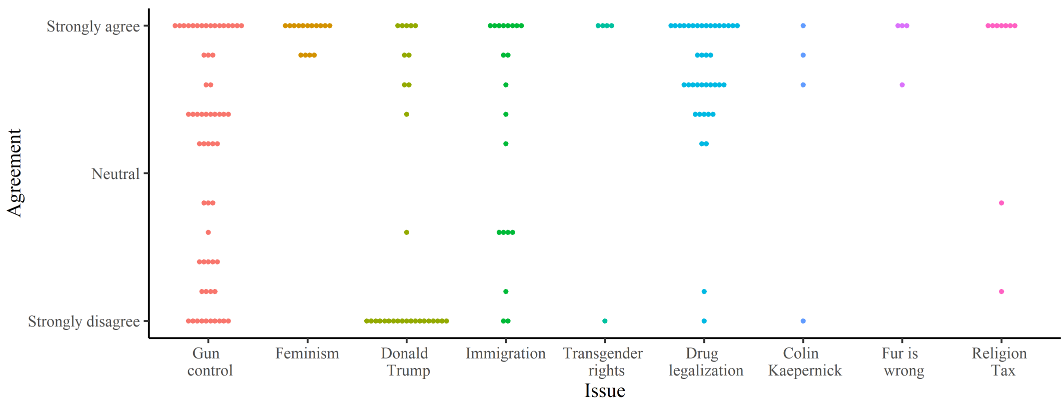

FIGURE D.2: Dot plot of the joint frequency of participants caring most about an issue and the extent to which they then agreed with the statement about that issue in Experiment 3 of Chapter 8.

Appendix E

Example Transcript from Chapter 8

The following is an example transcript of the instructions in Experiment 1 of Chapter 8 that could have been presented to a participant that indicated they cared strongly about gun control.

E.1 Instructions

We are following up on a previously published paper that looked at how people feel about moral dilemmas. In the previous paper, a moral dilemma was described that involved two possible courses of actions. Participants chose which action they preferred and had to rate how they would feel about performing that action. In this study, you will be presented with a scenario describing a moral dilemma. You will choose which action you would take and then provide a rating of how good or bad you imagine you would feel after taking that action.

E.2 Experimental trial

Imagine you have witnessed a man rob a bank. However, you then saw him do something unexpected with the money. He donated it all to a run-down orphanage that would benefit greatly from the money. You must decide whether to call the police and report the robber or do nothing and leave the robber alone. In the previous study:

- approximately 60% of participants who agreed with you about gun restrictions chose to call the police and report the robber.
- approximately 85% of participants who disagreed with you about gun restrictions chose to do nothing and leave the robber alone.

Would you:

- Definitely call the police and report the robber
- Very likely call the police and report the robber
- Probably call the police and report the robber

- Probably do nothing and leave the robber alone
- Very likely do nothing and leave the robber alone
- Definitely do nothing and leave the robber alone

E.3 Rating choice

You chose to call the police and report the robber. If you did call the police and report the robber, how would you expect to feel:

- Very good
- Moderately good
- Slightly good
- Neither good or bad
- Slightly bad
- Moderately bad
- Very bad

E.4 Understanding check

We were following up on a previous study in this task. Given what we described in the instructions, which of the following is true about the previous study?

- Participants chose which action they preferred (correct)
- Due to a computer error, participants were not allocated equally to imagine performing the different actions (incorrect)
- No data was saved during the experiment. (incorrect)
- The participants completed the experiment with their eyes closed. (incorrect)

E.5 Identity check

Please rate how much you agree or disagree with the following statements:

- I identify with Pro-Gun Enthusiasts
- I identify with Anti-Gun Advocates

Appendix F

Prior Analysis for Chapter 8

To inform the priors for the experiments presented in Chapter 8, we ran a Bayesian ordinal logistic regression on the data across Experiments 1A, 1B and 2 from Pryor et al., 2019b. The data from those experiments is available at the following repository: https://osf.io/n6uz5/?view_only=7dc67fcc0c1f4fdea8fe1dfe5d492480

We predicted responses to the moral dilemma as a function of whether the ingroup norm favoured reporting the robber or leaving the robber alone. We placed a weakly regularising prior on the effect of the ingroup norm (measured as a log odds ratio) using a normal distribution with a mean of 0 and a SD of 10. The posterior distribution for the effect of the ingroup norm had a mean of 1.27 and a SD of 0.24, and is shown in Figure F.1. We used this posterior to inform some of the priors for the current paper.

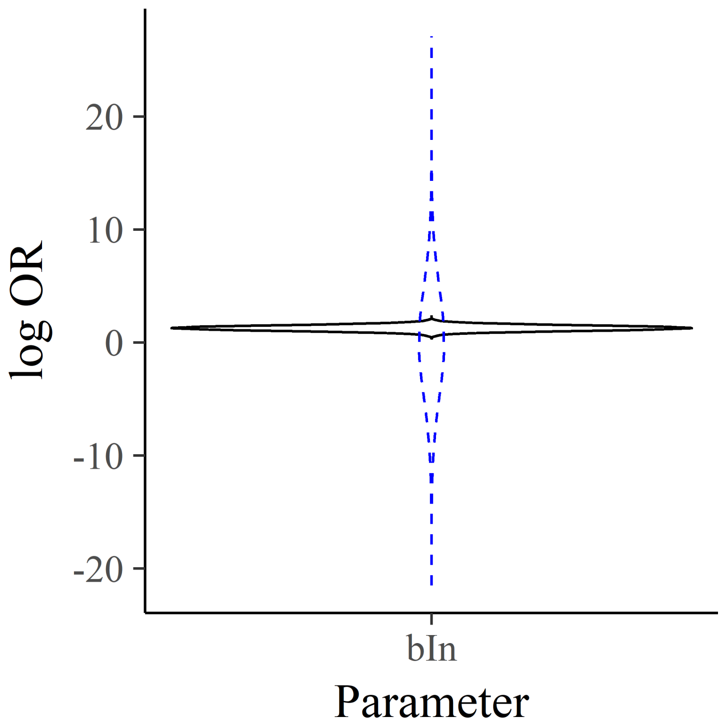

FIGURE F.1: Violin plot of the prior (dashed blue line) and posterior (solid black line) density for the effect of the ingroup norm in Pryor, Perfors, and Howe, 2019b.