# Effects of ‘Extralegal’ Factors on Adjudication

Paul Benjamin Troop
University College London

Submitted for the degree of Doctor of Philosophy (PhD)

# DECLARATION

I, Paul Benjamin Troop, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed …….

# ABSTRACT

Transparent, fair, and effective legal adjudication is integral to a healthy society, but adjudication does not always attain these aspirations. Miscarriages of justice and arbitrary outcomes undermine trust and confidence in the justice system. Addressing such challenges is problematic: legal adjudication is a complicated and oblique process, and our common sense or 'folk' theories of adjudication are often imperfect and not explicit or formalised. Legal psychology offers a more systematic way of providing insight, yet existing theories do not encompass or integrate all stages and aspects of the adjudicatory process, and struggle to account for some empirical observations. This thesis offers a more comprehensive theoretical account of legal adjudication that integrates key factors such as the values of the adjudicator, the applicable law, and reasons given for decisions. It suggests that some empirical observations that have previously been considered irrational behaviours can be integrated with such an account. The thesis provides empirical evidence for some of the central theoretical claims, demonstrating that adjudicators seem to take legally irrelevant or extralegal information into account to further their value outlook in a sophisticated and strategic way that is sensitive to the adjudicatory environment. The thesis also offers empirical evidence for asymmetrical order effects that arise when adjudicators consider pairs of similar cases sequentially, showing circumstances where adjudicators' later decisions are influenced to become more similar to earlier decisions, as well as circumstances where later decisions are influenced to be more dissimilar to earlier decisions. The thesis additionally indicates that casual explanations for offending behaviour based on genotype and previous childhood mistreatment can have a considerable mitigating effect on criminal justice outcomes.

# IMPACT STATEMENT

This thesis offers a more comprehensive descriptive psychological theory of legal adjudication in the context of a research field where there is a relative lack of such theories. The theoretical analysis integrates the diverse and heterogeneous stages of a trial with key and often overlooked factors that are characteristic of legal adjudication such as the law and reasons given by adjudicators. It also incorporates considerations such as the outlook and values of the adjudicator, enabling it to offer greater coverage and explanatory power. The theory offers predictions of some of the circumstances where legal adjudicators will depart from the norms of legal adjudication and take into account legally irrelevant or extralegal information. Whereas these behaviours have previously been deemed irrational, this thesis suggests that they have a rational foundation. These insights could be used to offer public policy recommendations to address these behaviours when they are considered unacceptable. The theoretical insights are also valuable to legal practitioners and litigants keen to understand, predict, and gain better outcomes from adjudicatory processes. The theory also generates hypotheses and predictions that could be further tested by academics in psychology, economics, law, and other disciplines to enable greater insights and more robust predictions.

The thesis also offers empirical evidence of the influence of values and extralegal information on outcomes consistent with the theory. This would facilitate more confident public policy prescriptions, as well as outlining areas for future research. Given that the empirical evidence is primarily from lay participants, the research provides testable hypotheses for other researchers to replicate and extend the findings using experienced decision makers and professional judges. The empirical research provides an overview of the state of knowledge in this area which also highlights areas of uncertainty which would benefit from future research. For example, there is an implication from the research that requiring adjudicators to provide reasons for their inferences at interim points may effectively moderate the influence of extralegal information on outcomes.

The thesis confirms the existence of order effects where adjudicators determine cases differently depending on the cases that they have previously determined. This also extends previous findings in that it also demonstrates order effects whereby subsequent rulings become less similar to prior rulings, a finding that is not well explained by current theories. This phenomenon is policy relevant in that it suggests that legal outcomes may be arbitrary in that they are path dependent rather than dependent on legally relevant factors. It also has academic impact given that the phenomenon requires theoretical explanation.

Findings that causal information about an offender's genotype and childhood abuse lead to strong mitigating effects in criminal justice contexts when risk is controlled for are directly relevant to legal practice given that they isolate the phenomenon with some precision. Furthermore, confirmation that the phenomenon does exist, contrary to some suggestions otherwise, reopens academic debates about why the phenomenon occurs, and also raises important public policy questions about whether this information should be admissible at all.

# ACKNOWLEDGEMENTS

I would like to thank my primary supervisor David Lagnado, who has supported me every step of the way with insight, patience, and humour. It has been a complete pleasure studying under his guidance. Adam Harris has also been a wise, careful, and genial collaborator throughout my studies. The Experimental Psychology Department in general and the Causal Cognition Lab in particular have been wonderful and supportive environments for both academic work and social life and places where I always felt welcome and valued. I am particularly grateful for Christos Bechlivanidis and Stephen Dewitt who were invariably available for sensible advice. Maarten Speekenbrink taught me almost everything I know about statistics and always responded with enlightening answers to queries. Henrik Singmann was hugely helpful in overcoming statistical problems I would have otherwise struggled with. Jeremy Skipper, graduate tutor, was always available, amiable, and astute and helped me successfully navigate my studies. On the administration side, John Draper and Antonietta Esposito's able assistance was also very important. John, George Joseph, and the rest of the Green Impact Team were a hugely meaningful part of my time at the university. It was a privilege and a pleasure working with Peter Howell, Katie Fisher, and the demonstrators I worked with while engaged as a head demonstrator. The Faculty of Laws was also a second home during my studies, not least because of the careful and judicious advice and support from my second supervisor Cheryl Thomas. Prince Saprai believed in and supported me while I was a teaching fellow in the faculty and was always available to answer questions about legal theory and teaching. A number of other Laws staff members assisted me in important ways, in particular Ben MacFarlane, Kevin Toh, Paul Burgess, Steven Vaughan, and Charles Mitchell. I am very grateful to the students who attended the Laws Work in Progress Forum where I presented various pieces of work. I am particularly grateful to all my colleagues at the Open University, but special thanks go to Paul Catley, Lisa Claydon, Anne Weismann, Simon Lee, Hugh McFaul, Emma Curryer, and Marjan Ajevski. I would also like to thank my two excellent examiners for this thesis, Lasana Harris and Yvonne McDermott Rees, who scrutinised my work carefully and thoroughly and provided many thoughts and insight to incorporate in my work going forward. Finally, I would like to thank Blanka who

was there at the inception and supported me throughout, Zoli and Mishi who arrived later and happily tolerated my commitment to my studies, and Keri whose respect and encouragement has been invaluable.

# TABLE OF CONTENTS

# TABLE OF FIGURES

# 1. INTRODUCTION

For those whose lives are touched by the criminal or civil justice systems, the way that adjudication operates matters very much. It is also intrinsically important to all of society that legal adjudication is transparent, fair, and operates the way that it is expected to. Yet we know from notorious miscarriages of justice and evidence of bias and arbitrary outcomes that it does not always proceed this way. In these circumstances, even if legal systems operate correctly, doubts and suspicions may remain. This thesis argues that psychology can be used to gain a better insight into these issues and how they might best be addressed. From a theoretical perspective it endeavours to integrate a number of different insights into a more comprehensive whole, arguing that some adjudicatory behaviours often labelled as irrational may have a rational foundation. From an empirical perspective, it aims to augment our understanding of adjudication by testing these theoretical assumptions and presenting evidence of the effect legally irrelevant information on outcomes; the effect of case order on outcomes; and the effect of causal explanations on criminal justice outcomes.

When addressing perceived shortcomings with adjudicatory systems, what can be called our common sense or folk theory of adjudication only takes us so far (Coleman & Leiter, 1993, p. 585; Leiter, 1997, p. 312). As we know, though common sense understanding is remarkable and has no real alternative in many circumstances, it does not always provide a complete or perfect model of legal adjudication. Common sense theories or models of adjudication also tend to be implicit rather than explicit, meaning that the assumptions of these theories may be absent or, if articulated, may not properly reflect the way common sense actually operates. Some have even suggested that common sense theories of behaviour rarely evolve at all (P. M. Churchland, 1981, p. 75). It is for these sorts of reasons that theorists from many different disciplines have historically looked for psychological theories of adjudication in place of, or to supplement, these common sense theories (Cohen, 1935, p. 821; Heller, 1979, p. 185; Kornhauser, 1984, p. 351; Leiter, 1997, pp. 271–272, 2001, pp.

282–283; Posner, 1987, pp. 778–779). Examples include the American legal realists of the early 20th century, many of whom were inspired by psychological behaviourism; research in the legal attitudinal tradition; the law and economics movement of the mid 20th century onwards that relies on rational choice theory; the more recent behavioural law and economics movement that questions the assumptions of rational choice theory using empirical observations; and the increasing popularity of empirical legal studies.

Given all this interest, there is a significant corpus of empirical psychological research into adjudication, but still a relative dearth of descriptive psychological theories of adjudication (Baum, 1997; Hirsch, 2003, p. 602 fn16; Posner, 2008, p. 19; D. Simon, 1998, p. 4, 2010, p. 131). One reason for this apparent deficit may be that adjudication is not a single homologous process, but a series of heterogeneous processes. As a result, many different psychological or psychologically inspired theories focus on discrete facets of adjudication, but few, if any, purport to provide a comprehensive account of all of adjudication. Another reason for the paucity of theories may be the additional complexity of integrating two matters that are particularly characteristic of legal adjudication: the law and reasons (Braman, 2009, p. 19; Knight, 2009, p. 1538; Rowland & Carp, 1996, p. 136). The law provides an, often incomplete, account of how adjudication should be undertaken, while reasons given by an adjudicator purport to represent how an adjudication has been undertaken. In some instances it is safe to assume that both bear a reliable relationship with reality, but both anecdotal and empirical evidence show that this is not always the case. Few psychological theories of adjudication address these two matters in significant detail.

Thus, the first task of this thesis, undertaken in Section 2, is to review the empirical and theoretical literature, take the existing disparate empirical and theoretical pieces of the existing puzzle, and assemble a more comprehensive theory. Given the relative paucity of psychological theories of adjudication, this relies on a more expansive survey of psychological theory, drawing on theories that do not strictly purport to be theories of adjudication but which nonetheless offer helpful insights. Examples include the story model of jury decision making and the law and economics movement. This more comprehensive

theoretical picture that results is presented in accordance with the rough chronology of a criminal trial and attempts to account for the different stages from opening speeches to final decision as well as how some of these stages can be integrated with one another.

Nonetheless, even this more comprehensive theoretical account is still demonstrably incomplete. While it accounts for a significant portion of legal adjudication, there remain empirically well evidenced instances of behaviour that are not well explained by the theory. For example, rather than legal inferences proceeding forwards, there are puzzling circumstances where inferences appear at least to proceed backwards. The leading psychological theories tend to put these unexpected departures from predicted adjudicatory behaviour down to irrationality, sometimes said to be caused by the complexity of the decision making task compared to the cognitive capacities of the decision maker, at other times said to be caused by decision makers slipping into using more holistic and heuristic 'System 1' cognitive processes rather than more logical and deductive 'System 2' cognitive processes. Sections 2 and 3 suggests an alternative theoretical explanation for these phenomena that is instead rooted in rationality. The suggestion is that while legal inferences appear to be proceeding backwards, what may actually be happening is that adjudicators are taking into account 'extralegal' or legally impermissible factors that matter to them and that ordinarily cannot be discerned outside the experimental environment, and then working backwards to present what will appear to standard observation to be a coherent and legally sustainable reasoning process that is compatible with their preferred outcome.

Section 4 turns from theoretical to empirical research, looking for evidence to distinguish between the irrational and rational explanations for the effects on adjudicatory outcomes that do not seem to fit with existing theory. Across a series of studies set in criminal and civil law contexts and involving straightforward and more complicated legal issues, single and dual manipulations are used to examine the effect of character on legal decisions where character is logically and legally irrelevant (or 'extralegal') to the issues being determined. The findings suggest that the extralegal effects are unlikely to be due to complexity or lack of adjudicatory cognitive capacity, and are instead more compatible with

quite a sophisticated and rational sensitivity to the adjudicatory environment when adjudicators are looking to secure outcomes they are more sympathetic to.

Section 5 looks at 'order effects' on outcomes, where adjudicators determine similar cases differently depending on the order in which they are presented. Asymmetric order effects have been found in the related research field of moral psychology, but are less well evidenced in the context of legal psychology. A number of studies using civil and criminal contexts are used to replicate these order effects in legal contexts. The findings also extend previous knowledge: whereas order effects in the moral context have shown responses that become more similar, in the legal context they can be shown to be more similar in some instances and less similar in other instances. The section suggests that these findings are not well explained by existing theories.

Section 6 examines a primarily empirical question of whether giving causal explanations for offending behaviour based on genotype and childhood abuse has mitigating effects on outcomes in criminal justice contexts. Despite increasing real world use, recent research has suggested that such information has little effect. The studies in this section examine whether there is in fact a 'double-edged' effect of this information such that any mitigating effects are cancelled out by the aggravating effects of increased risk. Two criminal justice contexts are used for this: parole where the primary consideration is risk and sentencing where other considerations in addition to risk are relevant. The results indicate that in the context of parole where risk is the main consideration, causal explanations have no discernible influence. However, in the parallel sentencing context, there is a strong mitigating effect of causal explanations provided the increased risk implied by these explanations is controlled for.

Section 7 provides a more detailed overview of the empirical findings and concludes the thesis.

# 2. THE THEORY OF ADJUDICATION

## 2.1 INTRODUCTION

Since at least the early 20th century, psychologists and lawyers have sought a more scientific theory of judicial adjudication (Bix, 2009, p. 190; Leiter, 1997, pp. 271–272; Martin, 1997, p. 10). From the 1920s, the American legal realists began to question mainstream formalist or legalist theories of adjudication, arguing that there was more to legal decision-making than the reasons that judges gave to explain their decisions, and that other factors such as human values influence trial outcomes in addition to the written law (Cohen, 1935, p. 812; Hart, 1961, p. 12; Leiter, 1999, p. 261; Llewellyn, 1931, p. 1241; *Lochner v New York*, 1905, p. 198 US 76; Oliphant, 1926, p. 228; Pound, 1910, p. 15). Despite the realist suspicions, neither their empirical skills nor psychological science that was at that time dominated by behaviourism were sufficient to construct an alternative psychological theory of adjudication (Hart, 1958, p. 606; Leiter, 1997, pp. 311–312; Posner, 1995, p. 393). This review examines the present state of the project of understanding the trial process in psychological terms with the benefit of present-day psychological research and theory and seeks to put the disparate pieces of theory together in a more comprehensive way.

Psychological science has advanced considerably since the early days of the 20th century when the behaviourism was in the ascendancy (Leiter, 1997, pp. 311–312, 2001, pp. 282–283; Posner, 1995, p. 393). Nonetheless, there are still relatively few descriptive psychological theories of judicial adjudication (Baum, 1997; Hirsch, 2003, p. 602 fn16; Posner, 2008, p. 19; D. Simon, 1998, pp. 4–6). Those theories that do exist do not purport to account for every aspect of the complicated processes of adjudication, so this review necessarily interprets psychological theories of adjudication broadly. For example, the story model purports to be a theory of jury decision making, but has elements that are also relevant to judicial adjudication. Another example is positive law and economics which merits inclusion because of its reliance on rational choice theory (Becker, 1993, p. 401; Bix, 2009, p. 204; Korobkin, 2000, p. 321, 2004; Korobkin & Ulen, 2000, p. 1055; Kysar, 2006, p. 115;

Posner, 1983, p. 1; Ulen, 2000, p. 795; Veljanovski, 2006, p. 49). Because existing theories do not always speak to every stage of the adjudicatory process, the primary organising structure of this review will be the general chronology of a trial. The various stages of the trial will be examined one-by-one and relevant psychological theory and findings will be introduced and integrated at the stages where they are relevant.

Taking a cue from the time of the genesis of attempts to build a psychological theory of adjudication, the review will be illustrated for pedagogical purposes by a fictional British trial, *R v Blackadder*, set in the early 20th Century, a court martial from a BBC historical comedy set in the First World War (Boden, 1989). This court martial serves to model a generic first-instance trial that includes both fact-finding and decision-making and that could be generalised to either criminal or civil context. For reasons of tractability, it will be assumed that there is only a single judge. Though the trial is obviously fictional, sufficient sense can be made of the facts to construct a plausible context. One of the key themes that will be highlighted by this review will be the suggestion that existing psychological theories of adjudication tend to underplay the influence of the adjudicator's human values.

## 2.2 THE PROTAGONISTS AND CHARGES

The key protagonist in the trial is the defendant, Captain Blackadder. He commands a unit deployed to the Western Front and faces two charges. The first, and most serious, alleges he disobeyed a lawful order to advance on the enemy (War Office, 1914, pp. 16–17). This would have been an offence contrary to s.9 of the Army (Annual) Act 1914, and carries a capital penalty. Over a series of days Blackadder's commanding officers sought to covey that Blackadder and the men under his command should advance on the enemy. But Captain Blackadder feigned communication problems such as crossed lines on the field telephone and a misaddressed telegram, to pretend that he had not received the order. His superiors then dispatched a carrier pigeon 'Speckled Jim' with the order to advance. Being extremely hungry, Captain Blackadder shot and ate the pigeon, giving rise to the second and less serious

charge, killing a carrier pigeon. This would have been an offence contrary to regulation 21A of the Defence of the Realm Regulations, punishable by a maximum sentence of 6 months' imprisonment.

The court martial is presided over by General Sir Anthony Melchett and a board of officers, though the latter feature little in the proceedings. To simplify the analysis, the influence of the members of the board will be set aside and the trial will be assumed to be decided by General Melchett alone. Of course, a fully fleshed out theory of judicial adjudication would need to explain the interaction, deliberation, and decision-making of groups of judges rather than just judges acting alone, and some tentative comments in this regard will be suggested at the end of this review. By the army delegating responsibility to General Melchett to determine guilt and, if appropriate, sentence, there is an expectation that General Melchett will try the case fairly according to the evidence. Where there is a conflict between General Melchett's views and those of the British state, there is a corresponding expectation that the latter should prevail.

## 2.3 THE PROSECUTION OPENING

Most adversarial trials begin with at least one speech. Characteristic of speeches is that they contain a concise summary or theory about what is said to have happened, and this will often be in the form of a story or chronological narrative (T. Anderson et al., 2005, pp. 321–322; Bex, 2011, p. 59). This story is used to explain, and often to persuade (T. Anderson et al., 2005, pp. 152, 155–156).

In *R v Captain Blackadder*, the charges previously outlined would be relatively uninformative without an associated story. So stories that the prosecutor, Captain Darling, might advance are that: (1) Captain Blackadder was afraid of being killed in combat, so he ignored orders to advance on the enemy by pretending he had not received them; and (2) Captain Blackadder was hungry and mutinous, so he ordered his orderly to kill and cook a

homing pigeon, which he then ate.

By contrast, while Captain Blackadder could offer an alternative story to explain away the evidence, he does not do so. This is a common approach in criminal trials where the burden is on the prosecution to prove the charges to the relatively high standard of 'beyond reasonable doubt'. In other circumstances, the accused may offer alternative stories such as that: (1) Captain Blackadder was keen to fight, but did not want to advance without explicit orders, which he never received, and (2) the homing pigeon was killed in crossfire, but given it was dead and Captain Blackadder was hungry, he saw no reason not to eat it. The prosecution and defence stories might diverge because of genuine but mistaken inferences from the evidence, or, as in the case of Captain Blackadder, through active misrepresentation motivated by the understandable human value of not wanting to be killed as a traitor. Similarly, the claimant and defendant in civil cases will offer, and the decision-maker may expect, stories to communicate their case at an early stage in the proceedings.

Though stories are pithy, they are a representation of a much richer underlying model (Devine, 2012, p. 220; Lagnado, 2021, p. 202). For example, though not set out explicitly, the story implies auxiliary hypotheses that the judge can predict once he hears the story. For example, the prosecution's story that Captain Blackadder is afraid and disobedient will imply that he would probably have demonstrated cowardice or disobedience on other occasions. By contrast, the defence's story will imply that there were probably other problems experienced with the communications between the central command and the trenches.

Story model theorists recognise that jurors sometimes compare stories offered by the parties (Bex, 2011, p. 81; Devine, 2012, p. 199; D. Simon, 1998, p. 29), but are not always clear on why the parties would volunteer such stories to the decision-maker at the outset. Often, stories are assumed to be created by the decision maker for their own comprehension and analysis (Hastie et al., 1983, pp. 22–23; N. Pennington & Hastie, 1991, p. 519, 1992, pp. 189–190). However, there are issues with this view. For one, stories are a highly condensed version of a much rich underlying model that is available to a decision-maker. As such, it is

not clear why a decision-maker would use an impoverished model when the full model is necessarily available to them. Secondly, stories have a very characteristic compressed narrative structure that makes them efficient to communicate an underlying model verbally, but this compressed and narrative structure seems to provide no obvious benefit when it comes to analysing a model. Thus it may be necessary to look beyond story model theories to understand the role of opening speeches.

The fact that a prosecutor or a claimant invariably provides a story at the very start of a trial suggests that stories have functions that are not just for the storyteller's own analysis. Anderson et al suggest efficiency and persuasion as two other functions for storytelling (T. Anderson et al., 2005, p. 157). Stories are representations of different models of what may account for the evidence. A multitude of models may be compatible with the evidence, with some being more and others less plausible. The lawyers for the parties analyse the material and isolate models within the hypothesis space that favour their respective clients. The judge can then assess these distilled models, which is more efficient than independently trying to replicate the processes already undertaken by the lawyers (T. Anderson et al., 2005, p. 157). In addition, the lawyers have an incentive to isolate the most plausible models as these are the ones most likely to persuade the decision maker (T. Anderson et al., 2005, pp. 152, 155). This suggests that stories articulated by the lawyers are used to share cognitive resources and to influence the cognition of the adjudicator.

Stories offered by lawyers at the outset of the trial will often be biased by their client's values. Parties who are genuinely in the right may have no need to finesse their stories, but parties in the wrong often will. The exception is, as Sperber points out, the odd occasion where truth is stranger than fiction. Here, a storyteller may instead be tempted to share a more plausible but untrue story (Sperber, 2001, p. 407). In Captain Blackadder's case, his lawyer will not simply expound the story that is just the most plausible, but rather the most plausible story that favours his client. As such, the client's human values, such as the desire to carry on living, colour the stories that their lawyers tell to the judge. This requires what Sperber and his collaborators term 'epistemic vigilance' on the part of the judge to avoid

being misled (Mercier & Sperber, 2009, p. 160, 2011, p. 60; Sperber et al., 2010; Sperber & Mercier, 2012, pp. 379–381). Epistemic vigilance is particularly pertinent to argumentation, and will be considered in more detail in the corresponding section below. Again, the fact that at least one important function of stories is to mislead other decision-makers challenges that story model assumption that stories are primarily for the story creator's own ratiocination.

A final issue to note is that model of what happened, in story or other form, is the outcome of other mental processes (Lagnado, 2021, p. 12). This raises challenging questions as to how these models are built in the first place, and how the judge assesses competing models. These are questions that we turn to next.

## 2.4 FACT-FINDING

The adjudicator's task, particularly at first instance, can be divided into two distinct processes, fact-finding (inferring the facts from the evidence) and decision-making (inferring the appropriate decision on the basis of the facts and applicable law) (Baron, 2008; Newell & Shanks, 2014, p. 19). Logically, the adjudicator must undertake fact-finding before decision-making. This is because regardless of the decision-maker's objectives, the achievement of these objectives always relies on an underlying accurate factual model. Crucially, fact-finding should be value-neutral because there is no logical basis for there to be an influence of what the factfinder wants on what he believes (Binmore, 2011, p. 5). As such, we should not expect fact-finding to be biased by values. Correspondingly, this means that there ought only to be a single correct factual model of what happened (Bayes & Price, 1763; Wigmore, 1913, p. 1).

As an independent tribunal of fact, General Melchett is not supposed to have witnessed events first-hand. He therefore needs to build a factual model by drawing inferences from the evidence presented at trial (D. Simon, 1998, p. 19). There are two different categories of evidence: real evidence covers things like the carcass of a roast pigeon

and feathers; and testimony is statements given by the soldiers who witnessed what happened. Both types of evidence may be equivocal and consistent with different stories, but particular care needs to be taken with testimony given the risk that this can be actively misrepresented (Mercier, 2010, p. 501; Sperber, 2001, p. 406). For example, Captain Blackadder is strongly motivated to avoid being shot as a traitor, so he seeks to misrepresent the evidence, telling those under his command to deny both receiving orders to advance and killing the homing pigeon.

To analyse fact-finding more deeply, it is helpful to distinguish between the different levels at which one can explain a cognitive process. Marr identified three different levels and pointed out that each embodies a different type of explanation (Chater et al., 2006, pp. 289–290; Marr, 2010).

Marr's first level looks at the goal and logic of the process (Marr, 2010, p. 25). In the context of judicial fact-finding, there is a reasonable degree of consensus at this level. In order to make a model of what happened, the adjudicator must draw inferences from the evidence (D. Simon, 1998, p. 19). This process consists of inferences because the information contained in the final model 'goes beyond the information given' in the evidence (Bruner, 1973; Mikhail, 2009, pp. 39–40; Todd & Gigerenzer, 1999, p. 10). The adjudicator has to rely on tacit pre-existing knowledge to generate new information from the evidence (Lagnado, 2021, pp. 118–119; Mercier & Sperber, 2011, p. 57; D. Simon, 1998, p. 42, 2004, pp. 520–521). This tacit information is sometimes called a generalisation or a warrant (Toulmin, 1958, p. 91; Walton, 2005, p. 15). For example, feathers are found on the floor of Captain Blackadder's bunker. Using the generalisation that feathers come from birds, General Meltchett can infer that these feathers must have come from a bird. Little or no learning takes place by the judge: the information that the judge's generalisations rely on has already been learned outside and prior to the trial. Some learning may take place at this stage, but in relatively discrete areas, such as where an expert witness is called to give evidence on domains outside the knowledge of the adjudicator. Quite how these generalisations are learned by an adjudicator remains relatively uncertain (Pearl, 2009, pp. 704–705), but sources

may include experiential, genetic, and cultural transmission (Lagnado, 2021, p. 25).

Given that inferences rely on generalisations that may vary between judges, different judges may come to different factual conclusions (Schum & Martin, 1982, p. 124; D. Simon, 1998, p. 75; Spellman, 2010, p. 149). For example, General Meltchett may have a very dim view of the obedience of the soldiers. This might lead him to be more inclined to conclude that Captain Blackadder had deliberately disobeyed orders and killed the carrier pigeon. By contrast, a judge with a more positive outlook of his soldiers might draw more favourable inferences.

Wider evidence can be integrated to draw even more illuminating inferences. There are complicated interrelations between pieces of evidence that often need to be taken into account. For example, there might be a number of hypotheses about the source of the feathers. The feathers might have come from an innocent source. But combined with the further evidence that the feathers are speckled and that there is the carcass of a pigeon on a plate on the table, taken altogether it seems probable that they in fact belonged to the distinctive missing homing pigeon 'Speckled Jim'. The process of simultaneously triangulating and integrating these disparate sources of information is sometimes called 'multiple parallel constraint satisfaction' and theorists have sought to model this process (Robbennolt et al., 2010, pp. 32–33; D. Simon, Snow, et al., 2004, p. 814; Spellman et al., 1993, p. 147; Thagard, 1989). Multiple parallel constraint satisfaction may be particularly valuable regarding witness testimony. Given that witnesses such as Captain Blackadder may be motivated by considerations other than a proper outcome, the adjudicator has to consider their testimony from a wider perspective. General Meltchett can triangulate Captain Blackadder's claims with the real evidence to look for inconsistencies (Mercier & Sperber, 2009, p. 160; Sperber et al., 2010, p. 375). Inconsistencies can be grounds to doubt the testimony (Engel, 2006, p. 250; Sperber, 2001, pp. 409–410; Walton, 2005, p. 48). To determine which factual model of what happened is most likely, adjudicators may therefore have to simultaneously consider much of the evidence.

Marr's second level of explanation is the algorithmic level. This examines how the informational inputs and outputs are represented, and what the algorithm for the inference is. This level of explanation coexists with the first level of analysis, but looks very different and hard to reconcile subjectively with a 'common sense' view of the process (Devine, 2012, p. 23; Horst, 2011, p. para 2.3; Posner, 2008, p. 11; Schauer & Spellman, 2017, p. 266; Sperber & Mercier, 2012, p. 369). Nonetheless, this level of explanation has a crucial role because it allows the theory to be formalised and tested (Chomsky, 1957, p. 5; M. Jones & Love, 2011, p. 171; Marr, 2010, p. 19; H. A. Simon, 1992, p. 153). If a theory fails when tested at this algorithmic level, then this suggests there is an issue with the theory, whereas if it succeeds at this level, neuroscientists can continue to examine at Marr's third, hardware or neuronal, level whether this algorithm is the one that is actually realised in the brain (Marr, 2010, p. 25).

In the legal fact-finding context, there have been attempts to specify the algorithm at the second level, for example by building more formal parallel constraint satisfaction models (Holyoak & Simon, 1999, p. 8; D. Simon, Snow, et al., 2004). These connectionist-type models have been partially inspired by third level assumptions about the workings of neurons in the human brain (D. Simon, 1998, pp. 81–82; D. Simon, Snow, et al., 2004, p. 816; Thagard, 2004, p. 243; but see Glymour, 1992;). Evidential inputs are connected to factual outputs through a series of intermediate connecting nodes. Each intermediate node can take a range of values so that its outputs influence subsequent nodes in different ways. As a result, the network can make different factual inferences depending on the evidential inputs. The whole network can be trained using real data to adjust the values of intermediate nodes based on known correct inferences (M. Jones & Love, 2011, p. 172; D. Simon, 2004, pp. 520–521; Spottswood, 2013, pp. 9–10). Ideally, the network will then make accurate inferences when presented with novel data. Thagard has previously created a connectionist network to undertake various inferences including legal inferences (Thagard, 2004). Some theorists, most notably Dan Simon, feel that these connectionist networks are a fair representation of the process by which judges take into account numerous inputs to derive their factual findings (D. Simon, 1998, pp. 81–82; D. Simon, Snow, et al., 2004, p. 816).

One of the most important tasks for a theory of judicial decision-making is to explain how an adjudicator either builds their own factual model of what took place, or prefers a particular factual model out of those presented to them by the parties, particularly because we know that almost all of this happens outside the scope of conscious introspection. Story theorists have identified various considerations they call 'certainty principles' that are likely to be relevant to how adjudicators undertake this process. For example, one principle, 'coverage' is how much of the evidence a story can account for. If the story accounts for all the evidence it is more likely, *ceteris paribus*, to be accepted than a story that does not account for all the evidence (N. Pennington & Hastie, 1991, pp. 527–528). Nonetheless, these somewhat heuristic descriptions of the process are not yet very well specified, a point that story model theorists concede (Lagnado, 2021, p. 202; N. Pennington & Hastie, 1991, p. 550). As a result, the story model is not yet sufficiently developed to be able to shed much light on the process of model construction or selection.

The development of formal connectionist and story models has nonetheless decelerated in the legal arena. The network used by Thagard was not trained in the way that connectionist networks are expected to be, rather the values were specified in advance. One reason for this may have been the sheer quantity of tacit information that human adjudicators rely upon in making factual inferences, and the corresponding difficulty in training a connectionist, or any other, formal system. As such, this research programme tells us a limited amount about legal inferences using connectionist networks because we already know that connectionist networks are capable of embodying a wide range of logical inferences (M. Jones & Love, 2011, p. 172). Connectionism more widely has also suffered criticism due to the oblique nature of the representations at the intermediary nodes, which makes their workings very difficult to unpick (Smolensky, 1988). Similarly, the limited formal specification of story models has limited their development (Lagnado, 2021, p. 203).

Another means of representing legal inferences at the second level is provided by Bayesian approaches (Bovens & Hartmann, 2004), and in particular Bayes nets (Dawid & Evett, 1997; Edwards, 1991; Fenton & Neil, 2013, p. 141; Kadane & Schum, 1996; Taroni et

al., 2014). Bayes nets similarly represent the inputs and outputs via intermediate nodes, but the intermediate nodes are more meaningful than connectionist models in that they represent recognisable aspects of the world (Fenton & Neil, 2013, pp. 119, 139; M. Jones & Love, 2011, p. 170). For example, in our Blackadder example, there might be a node representing the likelihood that the feathers came from a pigeon or another bird and another that represents the likelihood that the carcass came from a pigeon or another bird. Each input generates a probabilistic output based on various conditions. These conditions can either be calibrated by teaching the Bayes net using objective training data in a similar way that connectionist networks are trained, or it can be specified using subjective values provided by a researcher (Fenton & Neil, 2013, p. 34). As noted above, given the difficulties with the massive reliance by adjudicators on tacit information, and the fact that before training any node could theoretically be linked to any other node, the researcher tends to specify the relationships in advance. Nonetheless, the Bayes network approach is valuable in that its transparency allows the theoretical assumptions to be specified with some rigour.

To date, many theorists have discounted the value of Bayesian approaches, but the reasons for this are worth interrogating. For example, Pennington and Hastie dismiss a Bayesian approach on the basis that some participants gave inconsistent ratings on guilt and innocence (N. Pennington & Hastie, 1991, p. 549, 1992, pp. 199–201; Spottswood, 2013, pp. 6–7). Given that guilt and innocence are supposed to be different sides of the same coin, when ratings of one go up, ratings of the other should go down and vice-versa. Yet Pennington and Hastie found that almost half of participants' responses did not do so (N. Pennington & Hastie, 1992, pp. 199–200). This seems puzzling, but perhaps too hasty to discount Bayesianism given that the majority of participants did act in a way that the theory would predict, and the lack of a plausible explanation as to why participants were giving apparently inconsistent responses. Others such as Devine have rejected Bayesian approaches because they find them hard to reconcile with the intuitive picture of judicial proceedings (Devine, 2012, p. 23), but as noted above, a Bayesian approach fits at a different theoretical level of explanation and as such ought not to be expected to mesh with a common-sense level of explanation.

Dan Simon raises another objection to Bayesian approaches. He argues that some empirical evidence suggests that Bayes is not an appropriate model for legal inference because empirical evidence suggests legal inferences go 'backwards', contrary to the tenets of Bayesian theory (D. Simon, 2004, p. 514; D. Simon, Snow, et al., 2004, p. 814). For example, in experiments carried out by Simon and his collaborators, he examined the likelihood that a suspect had committed a theft. One of the pieces of evidence was an eyewitness. Participants indicated their assessment of the likelihood that that witness was correct. Thereafter, they were given evidence of either a positive or negative a DNA match with the suspect. After receiving the DNA evidence, their assessments of the correctness of the eyewitness identification changed so that a positive match increased their assessment that the eyewitness was accurate, and a negative match decreased their assessment that the eyewitness was accurate (D. Simon, Snow, et al., 2004, p. 822). Simon concludes that this pattern would be contrary to a Bayesian approach on the basis that Bayesianism theory assumes that inference is unidirectional (D. Simon, Snow, et al., 2004, p. 822). However, as has previously been pointed out by Lagnado and Gerstenberg, this is to misunderstand Bayesianism (Lagnado & Gerstenberg, 2017). Evidence that an eyewitness identification has been confirmed (or disconfirmed) by a DNA test increases (or decreases) the probability that the eyewitness was correct. It is therefore appropriate to adjust one's assessment of the eyewitness so this pattern does not cast doubt on Bayesianism.

But another empirical pattern identified by Simon seems to raise a more fundamental challenge to Bayesian approaches. Simon found that some evidence influenced the finding of facts where the evidence bore no plausible relationship with those facts. For example, Simon et al found that the character of a party seemed to influence the finding of whether the internet was more similar to a newspaper or a telephone network (Holyoak & Simon, 1999, p. 12). Essentially, this means that factual inferences appear to be influenced by values, a pattern that we have already noted ought not logically to happen (Binmore, 2011, p. 5). This empirical pattern has been repeatedly replicated (Liu, 2018, p. 96; Liu & Li, 2019, p. 637; Spamann & Klöhn, 2016, p. 255; Wistrich et al., 2015), and does at first blush cast doubt on

Bayesian approaches. However, returning to the theory, anomalous findings may instead be evidence of an problem with auxiliary hypotheses associated with the theory (P. S. Churchland, 1986, p. 261; Duhem, 1914; Hempel, 1966, p. 28; Quine, 1951, 1960). Accompanying the hypothesis that Bayesian inferences go forwards are the auxiliary hypotheses that: (1) neither an adjudicators' fact-finding nor his decision-making are influenced by impermissible non-legal matters such as character, and (2) the reasons that adjudicators' give are generally veracious (N. Pennington & Hastie, 1991, p. 531; D. Simon, Snow, et al., 2004, pp. 826–827; Zamir et al., 2014, p. 668). Simon concludes that there is an issue with the sufficiency of Bayesian approaches to represent judicial decision-making. Yet there is an alternative possibility that Bayesian approaches could be suitable for representing judicial behaviour, but the empirical pattern is caused by the effect of judicial values. An example would be if General Meltchett found Captain Blackadder guilty of the serious charge of disobeying an order and sentenced him to death, not because he was satisfied of his guilt, but because of his antipathy caused by the knowledge that Captain Blackadder had killed and eaten his pet pigeon. It is this last possibility that we will explore and pay some attention to in rest of this section and then return to in Section 3.

## 2.5 DECISION-MAKING

As previously noted, fact-finding is followed by decision-making. Whereas fact-finding is an inference that seeks to determine what happened in the past, decision-making is an inference that will have consequences in the future. For example, General Melchett could acquit Captain Blackadder or he could find him guilty. If he finds Captain Blackadder guilty, he can choose between a range of sentences, ranging from imprisonment to capital punishment.

In making a decision, the judge might bear in mind the consequences for wider society (Schauer, 2010; Sunstein & Ullmann-Margalit, 1999). The adjudicator's decision will of course have immediate consequences for the parties in that they may or may not be held

liable and required to provide a remedy, or they may be held culpable and punished. But the consequences of the decision may spread wider. If the principles that the adjudicator uses to determine the case are novel, this may affect the position of other members of society due to the principle of *stare decisis* (J. H. Baker, 2002, p. 199; Engel, 2006, p. 225; Posner, 2008, p. 155; Rowland & Carp, 1996, p. 154; Rubin, 2000, p. 549; M. Shapiro, 1972). *Stare decisis* obliges judges in to follow the same principles established in earlier cases where they are faced with similar circumstances. For instance, if General Meltchett imposes a heavy penalty for killing a carrier pigeon, and thereby establishes a principle that such offences attract heavy sentences, other courts may be expected to do the same. Civilians or soldiers who might otherwise have engaged in such behaviour may thus refrain from doing so. Thus, General Meltchett's decision may have instrumental characteristics in that it can be used to further the state's policy goals across society.

When an adjudicator makes a decision, he does so based on considerations that go well beyond those of fact-finding. As noted above, the key criterion for assessing the quality of factual inferences is accuracy because the capacity of a judge's decision to further any policy goal depends on factual accuracy. By contrast, the decision making stage may be much more subjective because it is also based on values (Epstein et al., 2013, p. 385; Karni, 2005; Oaksford & Chater, 1998, pp. 4–5). Whatever decision the adjudicator makes will have different consequences in the world, and those consequences will matter more or less to members of society including the adjudicator. The government is concerned to prevent the unnecessary killing of carrier pigeons. General Meltchett is incensed with the killing of Speckled Jim and he clearly prefers that the perpetrator is sentenced to death to being given a fine. Divergences between the values given to these considerations by different members of society cause important considerations for theories of adjudication.

While there may be many areas of law where judges' values will be very similar (Rachlinski et al., 2017, p. 2051; Sisk & Heise, 2004, p. 746; Sunstein et al., 2006, p. 48), we know from anecdote and empirical research that values in some areas of law are particularly polarised. Notable examples include attitudes to abortion and the death penalty (Lord et al.,

1979, p. 2098; Redding & Reppucci, 1999, p. 31; Sunstein et al., 2006, p. 55), political outlook (Epstein et al., 2013, pp. 77–78; Furgeson et al., 2008, p. 219; Maveety, 2003; Pinello, 1998, p. 219; Pritchett, 1941, p. 892, 1948; Rachlinski & Wistrich, 2017, p. 205; Schubert, 1962, 1965; Segal & Cover, 1989; Segal & Spaeth, 2002; Sisk & Heise, 2004, p. 746; Spaeth, 1961; Ulmer, 1960), gender (Peresie, 2004, p. 1761; Rachlinski & Wistrich, 2017, p. 207) and race (Cox & Miles, 2008, p. 1; Rachlinski & Wistrich, 2017, p. 207). Importantly, the evidence suggests that the characteristics of the judge do not indiscriminately affect outcomes. Rather, in order for characteristics of the judge to have an influence, it seems that the issues in the case also have to be personally salient to the adjudicator, for example gender issues to female judges (Peresie, 2004, p. 1761; Rachlinski & Wistrich, 2017, p. 207) and racial issues to minority judges (Cox & Miles, 2008, p. 1; Rachlinski & Wistrich, 2017, p. 207).

When making a decision, a judge takes into account a set of values and their imperfect factual model of the world to make a decision that they assess is likely change the state of the world so as to further those values. Given the integral importance of values to decision making, it is important at this point to distinguish the personal values of a judge from the more general values of a legal system. A judge will value consequences of legal decisions in a particular way. One way of thinking of the personal values of a judge is by considering how that judge would determine cases in the absence of other consequences, such as the obligation to follow previous precedent, or repercussions from other members of society. These personal values can be contrasted with the general values of a legal system: in areas where the law is unambiguous (which, as we shall see further below, is not always the case) it will be possible to infer these more general values. For example, where the criminal law places a burden and significant standard of proof on the prosecution, we might infer a higher general value of the system on avoiding miscarriages of justice than securing convictions. Similarly, where the civil law imposes strict liability for harm caused by particular activities, we might infer a higher general value of the system on avoiding the harms of those activities than on imposing liability for blameless behaviour. As we noted above, the evidence suggests that there are wide areas of law where the personal values of a

judge and the general values of a legal system coincide, but crucially for our analysis, there remain areas where the personal values of different judges diverge from the general values of the legal system (as well as from one another).

Thus, we can correspondingly draw a distinction between what can be called ‘legal’ and ‘extralegal’ information within judicial decision making. Where the law is sufficiently settled so that the values that it embodies can be unambiguously inferred, any information that evidences how those values will be furthered or undermined by a decision can be considered as legal. By contrast, where a judicial decision maker takes into account information that shows how any other values would be furthered or undermined by a decision, that information can then be considered extralegal.

In the so-called 'attitudinal tradition' of legal research, there is considerable evidence that adjudicators are influenced by their personal values (Epstein et al., 2013, pp. 77–78; Furgeson et al., 2008, p. 219; Pritchett, 1941, p. 892, 1948; Schubert, 1962, 1965; Segal & Cover, 1989; Segal & Spaeth, 1996b, 2002; Sheehan et al., 1992; Spaeth, 1961; Tate, 1981; Ulmer, 1960). Thus, General Meltchett believes that Captain Blackadder killed Speckled Jim and wants to impose capital punishment. By contrast, a more liberal judge would be inclined to impose a less serious penalty.

Just as we previously explored fact-finding at the second, algorithmic level, using Bayes nets, we can do the same for decision making (Posner, 2008, p. 11). Decision networks are similar to Bayes nets in that they have nodes connected by causal links, but in addition to the nodes representing probabilistic facts about events in the past, additional nodes represent decisions available to the adjudicator and the events that are expected to follow in the future from those decisions (C. L. Baker et al., 2011, p. 2470; Russell & Norvig, 2010, p. 626). Decision networks also include further nodes to represent the utility (or value) that the adjudicator would expect to secure depending on the decisions that he makes (Russell & Norvig, 2010, p. 627). Thus the utility node in the decision network representing General Meltchett's determination of the case would attach a high utility to finding Captain

Blackadder guilty and sentencing him to death, but a very low utility to acquitting him.

We can therefore say that where judges hear the same evidence, but make different final decisions, the variation could be the product of either or both of the fact-finding or decision-making stages. At the fact-finding stage, judges' factual inferences may vary due to the different generalisations that they draw from the evidence. At the decision-making stage, judges' decisions may vary due to their values or preferences. Thus either or both of generalisations or values may cause the variation in outcomes. However, judges are rarely entirely unconstrained in the inferences they make. Ordinarily there are constraints on their behaviour such as the law, which we will consider next.

## 2.6 THE LAW

At both the fact-finding and the decision-making stages, the law imposes restrictions on an adjudicator's inferences (Braman, 2009, p. 22; Brigham, 1978; George & Epstein, 1992, p. 323; Johnson, 1987, p. 325; O'Neill, 1981, p. 626; Segal, 1984, pp. 899–900; Sunstein et al., 2006, p. 82). No judge is an island. He is part of a justice system that implements his decisions. In a democracy, this system will encompass much of the population. If the judge behaves in a way that other members of the population find objectionable, they can act in ways to thwart his behaviour (Posner, 2008, pp. 87, 150, 156). In Blackadder, the Minister of War gets wind of Captain Blackadder's botched court martial via dispatches and intervenes to reverse General Meltchett's decision. A judge's decision may be overturned on appeal, a judge may be removed from his office, or parliament may legislate to remove powers from the judiciary as a whole. In some circumstances, those responsible for putting the General's decision into effect might engage in civil disobedience, and in an extreme situation, those who feel strongly about the inadequacies of the justice system might even attempt to overthrow it.

Law is often thought of principally as the 'black-letter' rules set out in parliamentary

statutes and common-law judgments (Posner, 2008, p. 252), but the limits on an adjudicator must necessarily be wider than this (Jeremy Bentham, 1780, pp. 303–304; Korobkin & Ulen, 2000, p. 1073; Sunstein, 1995). While nowadays much of the law is contained in black-letter sources, it was not always this way, and there remain areas where there is little black-letter guidance. In such 'cases of first impression', a judge must make a decision without his decision being determined by black-letter sources (Hirsch, 2003, p. 618). However, he is not free to impose any decision that he would like. For the reasons set out above, his decision is influenced by what other members of society would accept as appropriate (Posner, 2008, p. 235). Posner describes the scope for a judge to take a decision that is not going to be impugned as 'the zone of reasonableness' (Posner, 2008, pp. 86–87). If a judge in a case of first impression decided the case in a way that others found objectionable, the decision would still be likely to be overturned by one of the methods previously outlined. If the decision was reviewed by an appeal court, the rule embodied in the reasons of the appeal court would subsequently become black-letter law. Thus, returning to the idea of the general values of a legal sysem outlined in the previous section, any process of inferring these values would need to consider both the black letter law as well as how judges would determine cases in the circumstances where the black letter law is incomplete or uncertain.

Quite how tightly the law limits a judge is an important issue for psychological accounts of adjudication and also links into a longstanding debate within legal philosophy. Within the legal academy, a widely accepted view is that the black-letter law constrains fact-finding and decision-making in some, but not all cases (Braman, 2006, p. 310; George & Epstein, 1992, p. 323; Hart, 1961, pp. 124–154; Ho, 2008, p. 35; M. S. Moore, 1980, p. 161; Posner, 2008, p. 87; Schauer, 1988, p. 510; D. L. Shapiro, 1986, p. 737; Sloman, 1996, p. 11). There will be some laws applied to areas of ideological consensus where all judges will agree on their application ('easy cases') (D. Simon, 1998, pp. 19, 44). But the same law, applied to other circumstances, may lead reasonable judges to disagree ('hard cases') (Hart, 1961, pp. 12–13, 123). This position is consistent with the empirical psychological evidence that shows that judicial values have the greatest influence when the law is absent or contested (Braman, 2009, pp. 107–109; Braman & Nelson, 2007, p. 954; Johnson, 1987, p. 338;

Sunstein et al., 2006, p. 82; Wistrich et al., 2015, p. 900).

Related to the issue of whether the law imposes a tight or lose constraint on decision making is the issue of analogy drawing. Drawing an analogy is also a type of inference (D. Simon, 1998, pp. 19, 42). If a scenario is analogous to that of a previous case, then the rule in the previous case will apply. But if the scenarios are disanalogous, then the previous rule will not apply (Schauer & Spellman, 2017, p. 252). So similarly to how rules are applied, there will be easy cases where judges will agree that the scenarios are analogous, and hard cases where judges will disagree. While there is consensus that a theory of analogies is an essential element of a theory of adjudication, there is similar consensus that this is an underdeveloped area of research (Leiter, 1996, pp. 259–260, 271–272; Posner, 2008, p. 181).

It is noteworthy that the leading psychological theories of adjudication say relatively little about the impact of human values on judicial decision making. For Simon's psychological account of adjudication, the reason is deliberate. Simon focusses primarily on explaining the categories of cases where values are not salient to the judge. He writes: 'This model focuses on judging cases in which the judge is deemed to have no particularly important stake in either outcome. This is Hobbes' vision of a person divested of all fear, anger, hatred, love and compassion.' (D. Simon, 1998, p. 40). For Pennington and Hastie's version of story model the relative lack of focus on the influence of human values seems also to be the result of the scope of the theory combined with an underlying assumption that law imposes a relatively tight constraint on decision makers. Given that it is primarily a theory of jury decision-making, the story model does not purport to explain how the legal tribunal wrestles with questions of law, rather the focus is squarely on how the tribunal of fact applies the law that has been previously decided (Hastie et al., 1983, p. 22; N. Pennington & Hastie, 1991, p. 529). Nonetheless, even after legal issues have been determined, it is generally accepted, at least in the legal arena, that there remains some latitude as to how to apply the law to the facts. This more subtle picture contrasts with the story model's rather legal or formalist view of the law as more certain and implying decision making is essentially categorisation, a view that seems idealistic in the light of what we know about the nature of

law. The story model appears to assume that this latitude is relatively narrow because the wide variety of outcomes is put down to variation in generalisations relied upon at the fact-finding stage rather than the influence of values at the decision-making stage (Devine, 2012, p. 191; N. Pennington & Hastie, 1991, pp. 525, 534, 556; Spottswood, 2013, p. 2). Yet the empirical evidence suggests that values do have an influence, even in the cases within the story model's domain (Broeder, 1959, p. 748; Glöckner & Engel, 2013, p. 245; Kahan et al., 2012, p. 851; Wistrich et al., 2015, p. 900). Empirical research by lawyers, particularly in the attitudinal tradition, is also consistent with this (Posner, 2008, pp. 19–28). As a result, the story model seems to assume a somewhat overly simplistic view of adjudication and thereby places little or no attention on the motivation or scope that an adjudicator may have to further their own personal values as opposed to the general values of the legal system.

Finally, an observation that we will return to in the context of reason giving, is that due to the oblique nature of judicial cognition, legal rules can only limit adjudication if there is also some transparency about how judges reach their decisions (Fuller, 1978, p. 388; Liu, 2018, p. 83; A. Ross, 1946, pp. 65–66). From a simple judicial outcome alone, such as the fact General Meltchett found Captain Blackadder guilty and sentenced him to death, it can be very difficult for a external observer to infer the route followed by the judge through the evidence, generalisations, human values, and legal rules to arrive at that outcome. Many different routes will be compatible with the outcome, some of which some will be in accordance with the law and some will not.

## 2.7 ARGUMENTATION

Once the evidence has been heard, and before the judge makes a decision, the parties have an opportunity to present arguments to the judge in their closing submissions. Arguments have a different status to evidence. Arguments take the models set out in opening speeches as a starting point, but go further by additionally bolstering their own model and undermining that of the opposition (Walton, 2005, p. 1). General theories of argumentation

make reference to legal argumentation (Toulmin, 1958) and formally models of legal argumentation exist (Feteris, 2017; Prakken & Sartor, 2012; Walton, 2002).

Legal argumentation within the context of an adversarial trial has a number of characteristic features. The environment is asymmetrical in that the prosecution and defence seek to persuade the adjudicator General Melchett, but not vice versa. Persuasion also has to be sufficient not just to satisfy the adjudicator, but also to reassure the adjudicator that he can find reasons for his decision to satisfy others such as any appellate court (Feteris, 2017, p. xv).

Arguments themselves are reasons, or information, that support or undermine a particular inference (Mercier & Sperber, 2011, p. 57; Walton, 2005, p. 1). Arguments in a legal case may support or undermine a particular inference from evidence to facts or from facts to a decision. Distinctly legal arguments may reference legal rules that proscribe or prohibit certain factual or decision inferences (Feteris, 2017, p. 8). For example, in a criminal case, a rule against hearsay may forbid the adjudicator from making inferences based on hearsay evidence. A statutory or common-law rule may determine which factual circumstances amount to a crime. Argumentation consists of drawing the judge's attention to such information. Though every link in the chain of inferences will be supported or undermined by arguments, the lawyers' arguments will inevitably focus on the uncertain links where the judge might reasonably decide either way (Feteris, 2017, p. 3; Walton, 2005, p. 1).

Argumentation is valuable to the adjudicator because he will not always be fully cognisant of all the information (T. Anderson et al., 2005, p. 157). As with opening speeches, there is a particular value of argumentation where the case is cognitively demanding. Even a relatively modest trial such as *R v Blackadder* might entail a sizeable weight of evidence and a quantity of complicated law that a judge may struggle to fully marshal. As Spellman, Sperber and Mercier, and Fodor point out, individuals often do not realise the salience of information in their memory until it is pointed out to them (Fodor, 1983, p. 6; Mercier &

Sperber, 2009, pp. 154, 160–161, 2011, p. 60; Spellman, 2010, p. 155). General Meltchett wants to ensure that his decision and reasons meet a number of criteria such as being legally sustainable and which may also include furthering his value outlook. If his decision is unsustainable, an appeal court may overturn it, meaning his desire to see Captain Blackadder shot is thwarted. He may also sustain some damage to his judicial reputation (Higgins & Rubin, 1980, p. 130). In a complicated case there is a real risk that a judge might overlook the salience of a piece of evidence or a particular legal provision. There is therefore a benefit if the adjudicator takes advantage of the cognitive capacities of the lawyers for the parties whose arguments set out appropriate and inappropriate inferences based on the evidence and law that he may not have appreciated (T. Anderson et al., 2005, p. 157).

Argumentation is also valuable to the parties because it is an opportunity to influence the judge (Sperber, 2001, p. 404; Sperber et al., 2010, p. 360). There are two aspects to this. Where a party has evidence and law on their side, it is an opportunity to draw the judge's attention to this. But where the evidence or law is problematic, there will be pressure on a party to finesse the materials. This then requires the judge to exercise 'epistemic vigilance' to avoid being mislead (Mercier & Sperber, 2009, p. 160, 2011, p. 60; Sperber et al., 2010; Sperber & Mercier, 2012, pp. 379–381). Captain Blackadder refused to follow an order to advance and killed a homing pigeon to eat, but he has pleaded not guilty. Thus he has to present a plausible, but misleading, view of the evidence and law through his lawyer that is consistent with an acquittal. The adjudicator will therefore carefully scrutinise the respective submissions for inconsistencies that are the hallmark of a misleading view (Sperber, 2001, pp. 409–410; Sperber et al., 2010, p. 375; Walton, 2005, p. 48).

## 2.8 VERDICT

The empirical evidence confirms what American legal realists and many practising lawyers suspect: there are circumstances where adjudicatory outcomes are influenced by individual adjudicator's values. This is most prominent in research in the 'attitudinal'

tradition, where legal researchers have pointed out that a significant proportion of the variation in decisions at the appellate level can be explained by the judge's politics (Epstein et al., 2013, pp. 77–78; Furgeson et al., 2008, p. 219; Maveety, 2003; Pinello, 1998, p. 219; Pritchett, 1941, p. 892, 1948; Rachlinski & Wistrich, 2017, p. 205; Schubert, 1962, 1965; Segal & Cover, 1989; Segal & Spaeth, 2002; Sisk & Heise, 2004, p. 746; Spaeth, 1961; Ulmer, 1960), gender (Peresie, 2004, p. 1761; Rachlinski & Wistrich, 2017, p. 207) and race (Cox & Miles, 2008, p. 1; Rachlinski & Wistrich, 2017, p. 207). Consistent with theories of motivated reasoning, outcomes tend to be more strongly linked to the personal values of the judge where there is greater ambiguity in the steps that a judge has to take in making the decision (Boiney et al., 1997, p. 19; Braman, 2009, pp. 107–110; Braman & Nelson, 2007; Dana et al., 2007; Fried, 1996, p. 2142; Hsee, 1996, p. 122; Kunda, 1990, pp. 482–483; Segal & Spaeth, 1996a, p. 1075). At the same time, consistent with 'legalist' views of judging, law does impose something of a limitation on judges, but apparently not a complete one (Braman, 2006, p. 310; George & Epstein, 1992, p. 323; O'Neill, 1981, p. 626; Sunstein et al., 2006, p. 82; Wistrich et al., 2015, p. 900). If the evidence and law is unambiguous and there is little scope for a judge to choose outcomes, then personal values have limited or no effect. But where they are ambiguous, judges do seem to resolve that ambiguity consistent with their own values.

The influence of values often cannot be discerned in individual cases, it can only be shown as a statistical probability (W. M. Klein & Kunda, 1992, p. 146; Wistrich et al., 2015, p. 904). Thus, in the real world, it is very difficult to show from outcomes alone that a judge has been influenced by his personal values. Even on a statistical basis it is very challenging to show that an individual judge was so influenced. Rather, what can be said is that it is probable that a proportion of the judges were influenced without being a being able to identify which ones.

Some factors seem to consistently influence judges despite being legally irrelevant to the decision. The most striking example of such a factor is probably character. Where there is ambiguity in a decision and character is legally irrelevant to that decision, it nonetheless

seems to exert an influence on outcomes (Liu, 2018, p. 96; D. Simon, 2004, p. 538; Spamann & Klöhn, 2016, p. 255; Wistrich et al., 2015).

Other factors seem to affect outcomes in a more selective way. The evidence suggests that judges of a particular national, racial, or gender background do not consistently vote in a certain way in all cases. Rather, the issues in the case also have to be salient to that judge (Cox & Miles, 2008, p. 1; Peresie, 2004, p. 1761; Rachlinski & Wistrich, 2017, pp. 207–209). As a result, large areas of legal decision making are fairly uncontentious, with almost all judges of all backgrounds deciding in a similar way (Sunstein et al., 2006, p. 48). It is only in a limited range of cases, where particular issues are raised in a way that is salient to the adjudicator, that the pattern emerges.

### 2.8.1 Irrationality Based Explanations

A number of theorists put unexpected variation in final legal decisions down to irrationality. We noted above in the earlier section on the law that leading psychological theories of adjudication such as the story model and the psychological model downplay the effect of values, some because they focus on those cases where the judge has no interest in the outcome, and others because they do not focus on decisions involving ambiguity. As such, these theorists tend to put the recognised variation in outcomes down to differences in the generalisations that adjudicators rely on at the fact-finding stage or irrationality, rather than the influence of values at the decision-making stage (Hahn et al., 2015, p. 5; Hastie et al., 1983, p. 23; Kunda, 1990, p. 480; D. Simon, 1998, p. 75) or to simple error. For example, Pennington and Hastie write in the jury context: 'differences in story construction must arise from differences in world knowledge; that is, differences in experiences and beliefs about the social world.' and similarly: '[j]urors who construct different stories will either have brought different bases of world knowledge to the task or will have incompletely processed information presented at trial.' (N. Pennington & Hastie, 1991, pp. 525, 556). Likewise Dan Simon writes: 'This psychological model leads to the conclusion that, in the category of low-

stakes cases, we ought to overcome the surface similarity and reject the view that judicial opinions necessarily conceal result-driven judging.'(D. Simon, 1998, p. 136).

While it is highly likely that a proportion of the variation in outcomes is caused at the factual inference stage and also that errors occur, it seems unlikely that these factors completely account for the empirical patterns. For one, as we noted above, the key criterion for assessing factual inferences is accuracy, because achieving any normative goal at the decision stage depends on accurate factual inferences from evidence. But if factual inferences are being tainted by factors that are normatively irrelevant, this would seriously impair the quality of inferences. As Binmore points out, it is the equivalent of Aesop's fox inferring that chickens must be available because they taste better than grapes (Binmore, 2011, p. 5). Some theorists therefore bite the bullet and argue that adjudicator's inferences are indeed not in accordance with accepted tenets of rationality. Thus Dan Simon states: 'This account does not deny that human cognition is capable of performing some reasoning tasks in a manner consistent with the Rationalist account, nor that people violate these assumptions in some circumstances. Rather, it proposes a theory of cognition that contains elements of both approaches. A central tenet of the theory is that decisions are the product of a cognitive mechanism that operates bidirectionally, both in the prescribed and the reverse directions of reasoning.'(D. Simon, 2004, pp. 515–516; D. Simon, Snow, et al., 2004, p. 814). But there are challenges to this argument. For one, given the importance of accurate factual inferences from evidence, it is likely that there would have been strong selective pressures towards accuracy. Individuals whose inferences were biased by irrelevant and illogical factors would have been a disadvantage compared to those who were not. We could envisage a situation where environmental novelty may have caused evolved inference mechanisms to misfire outside their proper domain (Gigerenzer, 2000, p. 53; Kelman, 2011, p. 7), an idea we will explore in Section 6, but the types of inferences we are considering seem to be unexceptional. Simon has suggested that this assumed irrationality could be a product of the complexity of legal cases (D. Simon, 1998, p. 121, 2004, pp. 534–535), but these patterns manifest themselves even in fairly simple cases. Returning to the point about Bayesian inference noted within the earlier section on fact-finding: Simon's position requires him to argue that human

inference is not Bayesian, thereby challenging rational choice theory. But a more plausible explanation might be that human inference is Bayesian, provided the Bayesian account includes a judge's values in a utility function.

A second argument against irrationality theories is the directionality and nature of the inferences. If adjudicators factual inferences were simply irrational, or caused by lack of cognitive capacity, these inferences should be expected to be aimless and as likely to disfavour the adjudicator's values as to favour them. But the inferences are strongly directional, and moreover often in the direction that matches the interests of the adjudicator. Also, the patterns cannot be detected on an individual basis (Wistrich et al., 2015, p. 904), and if steps are taken to make the effects detectable on an individual basis, such as presented counterfactual cases side-by-side, the effect disappears (Nadler, 2012, p. 26; Sood & Darley, 2012, pp. 1343–1344). This does suggest that this behaviour is purposeful even if not necessarily conscious. As such, it seems somewhat likely that the effect is due to the human values of the adjudicator having an effect at the decision inference stage.

Dan Simon has a further reason for doubting that judges are influenced by their own personal values, which is that the empirical patterns manifest themselves as soon as the adjudicator begins to hear evidence, and even before they are aware that they are going to make a decision on that evidence (Furgeson et al., 2008, pp. 224–225; Holyoak & Simon, 1999; D. Simon, 2004, pp. 534–535, 540; D. Simon et al., 2001, p. 1250). Simon's view assumes that individuals who are not adjudicators do not manipulate information, and also that adjudicators would only manipulate information immediately before they make their decision. Neither assumption is necessarily safe. Outside the adjudicatory context, individuals can wield some influence on their environment by passing on information in a biased way so it might be unsurprising if they take an early opportunity to do so. Similarly, there would appear to be the advantages to an adjudicator manipulating information at an early stage in the trial. Where a judge sits as a panel, a biased model that is available early could be valuable to persuade their colleagues. Equally, bias would seem more detectable if a judge responded with an unbiased view during the course of evidence and only switched to a biased

view when they were asked for a decision as there would then be an obvious inconsistency between the two positions. Nonetheless, this remains an intriguing pattern that requires more definitive explanation.

### 2.8.2 Dual Process Theories

Other ways of explaining the empirical evidence of values on judicial outcomes rely on 'dual process' theories of cognition. Dual process theories assume that humans think using two different cognitive systems, often referred to as 'System 1' and 'System 2' (Haidt, 2001, p. 819; Kahneman, 2012, p. 21; Kelman, 2011, p. 33, 2011, p. 33; Korobkin, 2006, p. 56; Kysar, 2006, p. 109, 2006, p. 109; Newell & Shanks, 2014, p. 17; Sloman, 1996; Spottswood, 2013; Sunstein, 2005, p. 533). System 1 is assumed to be faster, automatic, parallel, and holistic. The output of its operation is said to be available to introspection, but not its operation or stages. By contrast, System 2 is assumed to be slower, sequential, deductive, controlled, and rule based, and its inference stages assumed to be available to introspection.

Some theorists are sympathetic to dual process based explanations of the influence of values on the judicial process because they take the view that such an influence is irrational (Liu, 2018, p. 84; Posner, 2008, pp. 107–108; Rachlinski & Wistrich, 2017, p. 223; Spottswood, 2013; Wistrich et al., 2015, pp. 863–864). Similarly, irrationality is sometimes said to be a hallmark of the putative System 1 (Mercier & Sperber, 2009, p. 149; Sunstein, 2005, p. 531). Unlike System 2, System 1 is thought to trade off accuracy for speed, relying on heuristics that do not take into account all the relevant information, thereby making System 1 prone to error (Kahneman, 2012, p. 25). Where a judge is influenced by factors that they are not supposed to be influenced by, these theorists see this as an error caused by the inappropriate use of System 1 inferences (Liu, 2018, pp. 88–89; Rachlinski & Wistrich, 2017, p. 223; Wistrich et al., 2015, pp. 863–864).

There are various issues with dual process theories of cognition that are beyond the

scope of this thesis. However, examples include the argument that the distinction between the two systems is hard to specify reliably (Sloman, 1996, p. 3), that the means by which the two systems interact is vague (Dhami & Thomson, 2012, p. 319; Kelman, 2011, pp. 35–36; Sloman, 1996, p. 3), that rule-based systems are as characteristic of parallel computations as they are of serial computations (Sloman, 1996, p. 5), and that it is difficult to predict theoretically which system will be engaged when (Gigerenzer, 2000, p. 291; Gigerenzer & Gaissmaier, 2011, p. 459; Lieder & Griffiths, 2019, p. 9; Posner, 1998, pp. 1560–1561; D. Simon, 1998, pp. 31–32). However, this review will confine itself to the issues that are particularly characteristic of legal adjudication.

A key question for dual process based explanations of adjudication is whether the apparent effects of human values are truly an error, or whether they are a rational response to a particular environment. The characteristic patterns identified by empirical research in the legal domain do not seem to be of the same type as other shortcomings identified by dual process accounts. Dual process theorists typically identify the shortcomings of System 1 processes as departures from the tenets of rational choice theory. But rational choice theory is agnostic about an agent's preferences (Binmore, 2011, p. 6). Anomalies might occur, for example, when an agent is inconsistent in pursuing their preferences. But from the perspective of the individual judge, it is tautologically rational to seek to realise their own objectives. It is also necessary to distinguish the real-life environment from the artificial environment created by experimental observation (Gigerenzer, 2000, p. 51; Kelman, 2011, p. 49; Pinker, 2009, p. 346). Behaviour that appears irrational with the benefit of experimental observation may be rational in a real life context where that behaviour is not as transparent. In the everyday environment of adjudication, it is often next to impossible to discern from an individual case whether a judge has been influenced by impermissible factors (W. M. Klein & Kunda, 1992, p. 146; Wistrich et al., 2015, p. 904). It is only in the controlled experimental circumstances where these patterns can be discerned, and even then this can only be done on a statistical basis. Therefore, it could arguably be rational for a judge to take account of their own personal values when such an influence would not be noticeable in the real world. Thus proponents of dual process theories may mischaracterise the relevant environment and

thereby misstate what would be rational (Kelman, 2011, p. 49; Todd & Gigerenzer, 1999, p. 13). If, given the correct environment, a behaviour would be rational, there is little more that a dual process account adds.

In addition, the dual process account does not seem to translate into the judicial decision making context terribly cleanly. Relatively little of the judicial inference process seems capable of meeting the criterion for a System 2 process of being available to introspection. When an adjudicator simultaneously considers the totality of the evidence in order to infer the most likely model of what happened, multiple considerations are weighed up in parallel in a way that the adjudicator can rarely articulate. The final model can be reported, but crucially not the process by which that model was created. Because this process is not accessible to introspection it does not seem to meet the criteria for a System 2 process.

Even once a final model is inferred, it is questionable how much access the adjudicator has to the individual generalisations supporting the inferences that contribute to that model. People have much less insight into their conscious processes than they assume (P. M. Churchland, 1981, p. 70; Dennett, 1984, p. 78; Hart & Honoré, 1985, pp. xxxiii–iv; Pylyshyn, 1999, p. 18; Schauer & Spellman, 2017, p. 266; Sloman, 1996, p. 6; Strawson, 1992, p. 7; Sunstein, 2005, p. 533). For instance, General Meltchett might infer from the evidence of feathers the fact that they came from a pigeon rather than, say, a chicken. But the generalisation that supports the inference (that a pigeon is more likely to be found in the trenches than a chicken) may not be as available to introspection as might be assumed (Mercier & Sperber, 2011, p. 58; Strawson, 1992, p. 5). It seems possible that the generalisation is actually inferred by the adjudicator rather than the adjudicator having direct conscious introspective access. For example, it seems relatively trivial for an adjudicator to infer the generalisation, given that it forms the necessary missing link between the evidence (feathers) and the conclusion (from a pigeon). It is therefore an open question whether the adjudicator has conscious and definitive access to the underlying generalisations or whether he works out them out on-the-fly as a necessary link between the evidence and the factual conclusion. In the absence of concrete cases, common law adjudicators certainly struggle to

articulate the generalisations underlying their intuitions (Goff, 1999, p. 318; Wambaugh, 1894, p. 56). Similarly, the generalisations that underly decisions seem equivalently elusive to articulation. Adjudicators, or any decision makers, often struggle to articulate these (Goff, 1999, p. 318; Kelman, 2011, p. 211; Mercier & Sperber, 2011, p. 59; Posner, 2007b, p. 437). In the context of the common law, initial decisions in response to a new problem are often treated with caution (Wambaugh, 1894, p. 56). It is only once a number of first instance and appellate level courts have examined the issue that consensus begins to form as to what the appropriate underlying principles are.

We also know that generalisations articulated in response to familiar problems often do not generalise well to novel problems. Generalisations applied to new contexts may lead to consequences that adjudicators find unpalatable. The most striking evidence of this is in the parallel ethical context of so-called trolley problems (Foot, 1967, pp. 8–9; Thomson, 1976, 1995). Theorists have tried to articulate the principles to account for different intuitions in the simplest toy dilemmas constrained to binary choices. Though it is relatively straightforward to differentiate between existing scenarios, theories are inevitably confounded when presented with novel scenarios that are only slightly different from the familiar ones (Appiah, 2008, pp. 90–91, 94–95). In the legal world, the articulation of principles for decisions is far from a straightforward process and often takes multiple iterations by different judges before a tolerably workable principle evolves. Principles are discerned, seemingly not because they are readily accessible to consciousness, but because they are worked out though trial-and-error.

In terms of the aspects of fact-finding and decision-making that are distinctly legal, the process by which analogies are drawn also seems not to meet the putative criteria for a System 2 process. The inputs: two factual scenarios, and the outputs: a view as to whether or not the scenarios are analogous, are all accessible to consciousness. However, the generalisations used to arrive at the assessment of analogousness have very limited conscious accessibility (Leiter, 1996, pp. 259–261; Posner, 2008, p. 12; Schauer & Spellman, 2017, pp. 254–258, 266). Again, the individual drawing the analogy may be able to articulate principles

that account for the difference, but as mentioned previously, it is not obvious that these principles are the operative ones rather than a plausible candidate generalisation inferred on the spot (Haidt, 2001, p. 822).

Similarly, the process of considering whether a law applies to given circumstances seems to be equally oblique. For example, regulation 21A of the Defence of the Realm Regulations provides:

'21A. If any person—

(a) without lawful authority or excuse kills, wounds, molests, or takes any carrier or homing pigeon not belonging to him ;

...

he shall be guilty of a summary offence against these regulations.'

We conclude instantly that Captain Blackadder's actions amount to an offence, including that shooting a carrier pigeon because he is hungry is not likely to amount to a lawful excuse. Yet the generalisations that underly the correct application of the key concepts such as 'kills' and 'lawful authority' are hidden (Mikhail, 2009, p. 36, 2011, pp. 19–20, 83). It is only with a range of known cases and considerable mental effort by lawyers that a tolerably workable but imperfect representation of the underlying generalisations can be worked out. Legal philosophers have sought to use the related technique of conceptual analysis which relies ultimately on introspection and intuition to tease out the underlying principles for the correct application of such legal concepts (Austin, 1956, pp. 13–14; Dworkin, 1977, p. vii; Grice, 1991, pp. 173–174; Kornhauser, 1984, p. 351; Leiter, 1999, p. 262, 2001, pp. 285–286). Nonetheless, it is generally recognised that even where undertaken by professional philosophers, the process is onerous and fallible, with many common legal concepts remaining contested (Leiter, 2003, pp. 49–50, 2012, para. 2; Stintzing, 1857, p. 107).

In addition to legal inferences not obviously being conscious, it is questionable the

extent to which they are the product of the application of deductive logic. When drawing a factual inference, an adjudicator does not seem to consider all evidence that might be available, nor every piece of knowledge that he possesses, or the possible interrelation between them. There are well known theoretical limitations such as the frame problem (Dennett, 1987; Fodor, 1987; Pylyshyn, 1987) and combinatorial explosion that preclude this (T. Anderson et al., 2005, p. 51; Engel, 2006, pp. 229–230; Gigerenzer, 2000, p. 227; Gigerenzer & Engel, 2006, p. 3; Pinker, 2009, p. 88; Spottswood, 2013, p. 7, 2013, p. 5). Instead, as the Bayesian nets discussed above illustrate, an adjudicator seems to consider a limited subset of the available information. As such, the inferences are characteristically heuristic, relying on known, and often statistical, generalisations (Saks & Kidd, 1980, pp. 126–127). Thus, it seems that the most plausible methods of undertaking many legal inferences in adjudication is to rely on methods that are characteristic of what dual system proponents would class as System 1 rather than System 2.

Finally, it is difficult to see how judicial inferences processes could meet the criteria of being 'controlled' if many of the moving parts appear not to be accessible to consciousness. An agent is not in control of a process if they are only in control of the parts that are accessible, with limited insight into the rest of the process. Thus, there remain many significant questions regarding dual process based explanations of some of the empirical patterns we find regarding adjudication.

## 2.9 REASONS

The final aspect of our survey of adjudicatory theories is that of reason giving. Here, relatively few psychological theories say much about the reasons judges give for their ultimate conclusions despite reason-giving being recognised in law as an important obligation. For example, Dan Simon speaks of reasons as a 'snapshot' of the judge's inference process, but does not analyse in much detail why a judge would give reasons (D. Simon, 1998, p. 35). Story model theorists similarly assume that in the jury context reasons

articulated to researchers are primarily veracious accounts of their thinking (N. Pennington & Hastie, 1991, p. 531). However, given that jurors generally do not provide reasons for their final conclusions, it is understandable that the theory behind reason giving is not very well developed in the story model context. Nonetheless, the evidence we have surveyed seems to suggest that the story model assumption of veracity is not always a safe one, particularly where the law is incomplete or uncertain and where adjudicators' values diverge from the general values of the legal system.

In contrast to theories that assume veracity of reasons, the influence of human values seem to count against such assumptions. Characteristic of reasons or stories is that they are communicated to others. There is a body of theoretical work that implies that unwavering veracity in communication is not likely to be an effective strategy for an agent, mainly because the interests of the teller and the listener do not always coincide (Mercier & Sperber, 2009, pp. 159–160; Sperber, 2001, pp. 405–407). Particularly where the risk of being uncovered is low, some degree of disassembly may be advantageous. For example, General Meltchett, who is keen to execute Captain Blackadder for killing his beloved pigeon, might be motivated to find the latter guilty of the more serious offence of disobeying orders to achieve this. But if General Meltchett said openly that this was what he was doing, his decision would be immediately overturned (Higgins & Rubin, 1980, p. 130). A more effective strategy would be to dissemble. Given we unhesitatingly accept that witnesses follow such a strategy, it ought not to be a surprise if some judges did it too. As such, story or psychological theory assumptions of veracity are unlikely to be reasonable in all circumstances.

Another explanation for the function of reasons is given by law and economics. These theorists model judicial behaviour using rational choice theory (Cooter & Ulen, 2012, p. 3; Kysar, 2006, p. 115; Posner, 2007a; Veljanovski, 2006, p. 21). Law and economics encompasses different schools, but characteristic is the assumption that judicial reasons are primarily instrumental (Cooter & Ulen, 2012, p. 414; Kornhauser, 1984, pp. 353–354; Korobkin, 2000, pp. 319–320; Posner, 1995, p. 403). Many legal economists believe that the

rationale for judges giving reasons is forward looking, namely to influence the behaviour of individuals in society who will find themselves in an analogous situation to the litigants in the immediate dispute. As such, judicial reasons are akin to a type of legislation (Posner, 2008, p. 81). The rules that judges expound impose costs on the behaviour of individuals in society (Hertwig, 2006, p. 393), and thereby influence their behaviour in accordance with rational choice theory (Korobkin, 2000, p. 321; Veljanovski, 2006, pp. 45–46). While a relative minority of proponents of law and economics recognise that judges sometimes pursue their own values when creating law, most accept the instrumental view of law (Kornhauser, 1984, p. 356, 2014).

The instrumental view of law must be partially correct. Due to the importance attached to the principle of *stare decisis* (J. H. Baker, 2002, p. 199; Engel, 2006, p. 225; Rubin, 2000, p. 549), members of society can place some reliance on the fact that if the situation that they find themselves in is analogous to one that has already been determined by the courts, the court will treat them similarly. As noted above, black-letter law will impose something of a limitation on subsequent judges. Celebrated cases such as *Roe v Wade* 410 US 113 (1973) clearly establish the lawfulness of behaviour in specific domains.

Yet the instrumental view does not seem to exhaust all the functions of judicial reasons. On the one hand, it has been pointed out that the instrumental assumption relies on a 'bewilderingly difficult' assumption as to how individual members of the public are influenced by judicial reasons (Hertwig, 2006, p. 394). Judicial reasons are not systemised, save in specialised texts and proprietary systems that the layperson does not have access to. Understanding the content of the black-letter law is therefore difficult without access to a lawyer (Ellickson, 1986; Engel, 2008, p. 275). Furthermore, some choices are difficult to predict in advance, minimising the scope to consult with a lawyer beforehand. Such considerations reduce the opportunities for law to play an instrumental function.

Jurists, by contrast, have suggested that there is also a 'backward looking' function to judicial reasons in that reasons are assumed to legitimise the decision (Knight, 2009, pp.

1543–1544; Posner, 2008, pp. 110–111; Schauer, 1995; D. Simon, 1998, pp. 12–13). This is consistent with the ubiquitous obligation to give reasons, even in trivial cases that do not establish a new principle. Thus, first-instance judges are obliged to give reasons for their fact-finding, even though it is rare that a factual finding would have an instrumental function. Instead, a legitimising function might help to explain judicial reason giving in a way that can be integrated with the influence of values. We noted above when analysing the effect of the law that legal adherence cannot be ensured without also having some mechanism to promote transparency in the inference processes that judges follow. If a judge does not give reasons for their conclusion, it is very difficult for an observer or a party to work out what their inference procedures were, and correspondingly whether these inference processes were compliant with the law. This is because multiple routes of inference may be compatible with the evidence and conclusion. By contrast, reasons can legitimise a decision because if a judge articulates their inference processes in reasons, an observer can scrutinise whether the inference process was compatible with the law much more readily (Feteris, 2017, pp. 18–19; Knight, 2009, pp. 1543–1544; Posner, 2008, pp. 110–111; D. Simon, 1998, pp. 14–15). An observer who is unfamiliar with the evidence can assume that the evidence as recorded by the judge is correct and scrutinise the inference process for internal coherence. An observer with a partial knowledge of the evidence can check for external coherence between the evidence recorded by the judge in his reasons and the evidence that the observer recalls was presented. In essence, the observer undertakes the same coherence checking process vis-a-vis the judge as the judge undertakes in relation to witness testimony. Consistent with this, Liu found that judges who were required to write reasons for their decisions were less susceptible to the influence of inappropriate values than judges who were not required to write reasons (Liu, 2018, p. 83).

Nonetheless, it seems that even when reasons are given, this still does not provide full transparency. For instance, evidence of a conflict of interest on the part of the decision-maker may be sufficient to overturn a decision, even if the reasons are, at face value, impeccable. Further, as we have noted in above in the context of decision-making, it can be very difficult for judges to provide complete insight into their own thinking. Another reason for this may

be linguistic: jurists such as HLA Hart have pointed out that some ideas are very difficult to represent in language (Hart, 1961, p. 126). As such, reasons may only provide an imperfect means of making judicial behaviour transparent (Posner, 2008, pp. 110–111). Because reason giving is imperfect, it leaves scope for the strategic influence of the types of values that have been detected through empirical research. There is empirical evidence that seems consistent with this view. For example, judges remain influenced by their values even where they have to provide reasons, for there is a statistically significant relationship between the values held by the judge and the outcome of the proceedings. Yet these prohibited influences are, unsurprisingly, rarely referred to in their reasons (Braman, 2006, p. 310; Liu, 2018, p. 96).

## 2.10 CONCLUSIONS

As this survey of theories of adjudication has shown, a complete psychological picture of the various stages of a trial remains quite incomplete. While some psychological theories of adjudication make quite specific claims about particular aspects of the trial, they are often quiet or silent on others. In addition, a key theme is that there are some quite anomalous empirical findings where adjudicators' decisions seem to be influenced by their personal values. It is such counter-intuitive findings that call most strongly for an explanation. But as we have seen, the two leading psychological theories of adjudication that speak most closely to these anomalies - Simon's psychological theory and dual-process theories – do not seem to provide a complete explanation of why they occur.

By contrast, focussing on the influence of judicial values and following this thread in and out of the different stages of a trial may provide illumination that goes somewhat beyond existing theories. Developing the common and readily recognised conflict between a party or a witness and the tribunal provides an archetype for the conflict between the tribunal and an appeal court and the public. Whereas the more predictable types of court proceedings will reflect the situation where a party simply has to tell the truth to the tribunal, the more anomalous proceedings will reflect the situation where the truth will not help a party. In the

latter category of cases, where there is sufficient at stake (as in Captain Blackadder's situation), there may then reason for a party to consider misleading the tribunal. Likewise, it is those cases where the values of the adjudicator conflict with the law where there may be more reason for the adjudicator to try to achieve an outcome that is influenced by non-legal matters. But just as the evidence against a party may be compelling or equivocal, it is only certain environments that are sufficiently ambiguous to allow an adjudicator to further their personal, non-legal, values. Nonetheless, there are, as we have seen, features such as the requirement to provide reasons for their decision that could provide a mechanism to discourage adjudicators from doing so. Reason-giving may discourage non-legal influences by making verdicts and outcomes more transparent. Just as the truth of testimony is assessed by examining the internal coherence of what a witness says and the external coherence of what they say and the evidence, the lawfulness of an adjudicator's decision is assessed similarly by looking at the internal and external coherence of their reasons.

Overall, there remains much to be said about the circumstances within these anomalous circumstances about issues such as when a particular adjudicator will, or will not, be tempted to further their own non-legal values. It seems plausible that this will depend to some extent on what is at stake: an adjudicator is likely to feel the pull of fundamental values more strongly than trivial ones. Equally, much may depend on the individual character of the adjudicator and on the nature of the audience. But though much remains to be filled in, values seem an important thread to weave into the tapestry of a psychology of adjudication. In the next section, we will explore these anomalous findings in more detail and suggest a positive theory that might account for them in rational terms that we can subsequently scrutinise experimentally.

# 3. A DEFENCE OF JUDICIAL RATIONALITY

## 3.1 OPENING

In the metaphorical dock is the rationality of legal adjudicators themselves. This paradoxical situation arises because leading psychological theories of legal reasoning allege that some quite counter-intuitive empirical findings are evidence that adjudicators are acting irrationally. The paradigmatic understanding of legal adjudication is that inferences should proceed forwards from evidence to facts and from facts to decision, constrained at both stages by the applicable law (D. Simon, 2004, p. 514), and ignoring any evidence, facts, or values that are legally irrelevant or 'extralegal' (Posner, 2008, pp. 70, 253). But legal psychologists have suggested that there are circumstances where inference appears to proceed backwards (Holyoak & Simon, 1999, p. 23; D. Simon, 1998, 2004; D. Simon, Snow, et al., 2004, p. 822; Zamir et al., 2014, p. 675), from decision to facts or from facts to evidence, akin to Aesop's fox inferring that the grapes were sour because he could not reach them (Binmore, 2011, p. 5) or an adjudicator inferring that a suspect's fingerprints must be on the weapon because the adjudicator wants to find the suspect guilty (Holyoak & Simon, 1999, p. 12).

The nature of the empirical evidence tends to be of the following type: an adjudicator has a decision to make in favour of one party or another, or in favour or against an accused. The final decision rests on one or more issues that are uncertain and could reasonably be determined either way. An illustration might be rules that apply differently to vehicles, with the facts of the case concerning an aeroplane, with the adjudicator being required to determine whether the aeroplane fits this category (Hart, 1958, pp. 606–607). There is then some extralegal information that is both logically and legally irrelevant to that issue, such as a party's excellent or abysmal character. In these circumstances, character ought to have no effect on the determination of that issue or, in turn, the final decision (Posner, 2008, p. 253). Yet there are circumstances where the bad character of a party seems to influence both the final decision and the determination of the issue (Braman, 2009; Braman & Nelson, 2007; Epstein et al., 2013, pp. 65–99; Holyoak & Simon, 1999; Liu & Li, 2019, p. 630; D. Simon,

Snow, et al., 2004; see Sood, 2013 for a review; Spamann & Klöhn, 2016; Wistrich et al., 2015; Zamir et al., 2014, p. 675).

As noted above in Section 2, leading psychological accounts treat this behaviour as irrational, putting it down to a failure of cognition caused by the limitations of an adjudicators' cognitive capacity compared to the complexity of the tasks that they face (D. Simon, 1998, p. 121, 2010, pp. 140–142). Some theorists make a related argument by appealing to 'dual process' theories of cognition that assume that people rely on two ways of thinking: 'System 1' an automatic, fast, unconscious, associative yet unreliable system, or 'System 2' a controlled, slow, conscious, logical and reliable system (Haidt, 2001, p. 819; Mercier & Sperber, 2009; Sloman, 1996; Sperber & Mercier, 2012, p. 370; Spottswood, 2013, p. 2). When the phenomenon described above manifests itself, the adjudicator is said to be using System 1, rather than System 2, thinking (Liu, 2018, p. 84; Posner, 2008, pp. 107–108; Wistrich et al., 2015, pp. 863–864).

This section explores the rationality of adjudicators, arguing that it may be possible to explain some of this behaviour in rational terms. Instead of amounting to a failing, this behaviour may be a feature, designed to realise the goals of an adjudicator in the normal legal decision-making environment. As such, the behaviour would include two related elements: the primary element is that adjudicators take advantage of the ambiguity of legal decision environments to use extralegal information to choose outcomes that they subjectively prefer. In a normal legal environment, this is very difficult for observers to discern. The secondary element to this behaviour is that where adjudicators are required to provide reasons for their decisions, they actively manage and manipulate this information to make it more difficult for observers to detect the primary behaviour.

## 3.2 EVIDENCE

### 3.2.1 Introduction

As we have seen, a notable empirical pattern in the psychology of adjudication is that adjudicators sometimes seem to take decisions based on information that is apparently legally irrelevant to the task and they rarely disclose this information in their reasons (Braman, 2006, p. 310; Liu, 2018, p. 96). While much of this research is undertaken using lay participants, professional judges also seem to exhibit the same behaviour (Liu, 2018; Liu & Li, 2019; Spamann & Klöhn, 2016; Wistrich et al., 2004). Generally, such patterns can only be discerned on a statistical basis where a sufficiently large number of adjudicators' decisions are analysed. As a corollary of this, it tends not to be possible to discern which individual adjudicators were so influenced, just that a certain proportion must have been (E. E. Jones & Davis, 1965, p. 225; H. H. Kelley, 1973, p. 108, 1987, p. 10; Snyder et al., 1979, p. 2298; Wistrich et al., 2015, p. 904). As noted above, these patterns seem to arise in situations of ambiguity where a case outcome is determined by issues that could reasonably be determined in either direction. Collateral information that is legally irrelevant to those issues such as character is then manipulated between two conditions, so that one group of adjudicators determines the case in the light of the information that one of the parties is of, say, good character, whereas the other group sees the same information save that the relevant party is now of bad character. Results suggest that statistically significant proportion of the judges in the good character condition determine the issues and the case in favour of the party of good character compared to those in the bad character condition. This pattern has many similarities with the more general phenomenon of 'motivated reasoning' in psychology where judgements and decisions may be influenced by the values of the decision-maker despite these values being apparently irrelevant to the task (Ditto et al., 2009, p. 310; W. M. Klein & Kunda, 1992; Kunda, 1990).

### 3.2.2 Empirical Overview

A legal example of the effect of extralegal information on an issue is given by Holyoak and Simon (1999, p. 21). Participants were presented with a case with a number of issues to determine. One of the issues was an analogy with a previous precedent, which was whether the internet was more analogous to a telephone or a newspaper. Preceding experiments with no character manipulation showed that verdicts were approximately evenly divided between both parties. The introduction of a character manipulation had a dramatic effect on verdicts, with 72% of participants subsequently finding in favour of one of the parties when he was of good character, but only 22% when he was of bad character. Crucially for present purposes, assessments of the internet analogy that was logically unrelated to character also changed in the same direction, though the effect did not quite reach statistical significance on standard levels ($p = .12$). Other issues also changed consistently with the verdicts, but these other issues were not as obviously unrelated to the issue of character.

Wistrich et al similarly gave professional US judges ambiguous mock cases to determine as part of judicial training (Wistrich et al., 2015). In contrast to Holyoak and Simon's research, Wistrich et al gave the judges fewer issues to determine per case, but the issues were similarly unrelated to character. For example, the judges were asked whether pasting a false entry visa into a genuine passport amounted to forging an identity card where the (legally irrelevant) background conditions were that the accused was entering the US illegally to track down an individual said to have stolen drugs from a cartel (bad character) or an individual entering illegally to earn funds to pay for his daughter's liver transplant (good character). 60% of judges in the bad character condition deemed the facts to amount to forgery, whereas only 44% of judges in the good character condition did so. The difference was statistically significant at a standard level (0.05). Another case involved an arrest for personal marijuana possession where the fictional statute provided a defence if a medical professional stated that there were therapeutic benefits to medical marijuana use. The issue was whether a medical statement needed to have been provided before the actual arrest. Where the defendant was of good character, 84% of judges determined the legal issue in his

favour, whereas only 54% of judges determined the issue in his favour when he was of bad character. Again, this was statistically significant. A third case was a constitutional challenge to a blanket strip search policy when prisoners were first brought into custody. The relevant four conditions were a combination of whether the prisoner bringing the challenge was of good or bad character and whether they were bringing the challenge as an individual or as part of a class action. Where the plaintiff was suing as an individual, 84% of judges granted the motion where the plaintiff was of good character, but only 50% where the plaintiff was of bad character. This too reached statistical significance. The effect was less pronounced where the plaintiff was suing as part of a class, with 65% of judges granting the motion in the good character condition and 51% granting in the bad character condition. In a fourth experiment, Wistrich et al asked bankruptcy judges to determine whether an accused was fraudulent in running up credit card debts when the individual did not have the means to pay off the debts. In one condition the individual ran up the debts to go on holiday, in the other the individual ran them up helping their mother who was battling cancer. Again, only 32% of judges determined the issue in the individual's favour when they were selfishly motivated, whereas 52% determined the issue in their favour when they were benevolently motivated, and again this was significant. A fifth case raised the Fourth Amendment issue that the seriousness of the offence is not considered when assessing the reasonableness of search and seizure. Judges were asked to assess the reasonableness of a random drugs test that prompted a subsequent search of the individual's locker where further drugs were found. In one condition the individual tested positive for marijuana and unsmoked marijuana was found in his locker, in the other, the individual tested positive for heroin and heroin was found in his locker. Notwithstanding the Fourth Amendment principle, only 44% of judges admitted the problematic evidence in the less serious condition, whereas 55% admitted it in the more serious condition, a marginally significant statistical result. In a sixth experiment, Wistrich et al found that judges imposed somewhat higher punitive damages awards where the defendant was said to be from another state rather than the same state.

In studies of Chinese judges, Liu and Li (2019, p. 637) provided professional judges slightly more complex scenarios with more than one issue. Again, character was irrelevant to

the issue to be determined. In a first case, the petitioner sought damages for breach of contract and the defendant sought to reduce the amount of contractually agreed damages on the basis that the law permitted this where the agreed damages was 'excessively greater' than the actual damages incurred. Here the character manipulation was whether the defendant was conducting an extra-marital affair with a government official or whether no character details were provided. Whereas none of the judges thought that the plaintiff should win the case where no character details were provided, 38.5% decided that the plaintiff should win when the defendant was of bad character, a statistically significant finding at the 0.05 level. In addition to outcomes being influenced by character, the interpretation of issues that would favour with the relevant outcome were also influence by character in the same direction. Thus 87.5% of judges thought that the liquidated damages were excessively greater (a finding favouring the defendant) when no adverse character information about the defendant was provided, but only 30.8% when adverse character information was provided. Notably, no judges referred to the effect of character in their reasons, despite being given the opportunity to do so. In a second experiment, an accused was charged with illegally breeding parrots in captivity. Issues included whether a species bred in captivity could be classed as a 'wild animal' and whether the accused was aware of this status. The character manipulation described the accused as a gambler or a good father. As before fewer judges (37.1%) thought that the accused should be convicted when he was of good character compared to when he was of bad character (64.9%), a statistically significant difference. Correspondingly, more judges considered that the accused was aware of the status of the parrot where he was of bad character, but there was no difference by character concerning whether a domestically bred animal could be considered a 'wild' animal. Consistent with previous findings, only a single judge referred to character in their reasons. In a third study based on a personal injury case, the plaintiff claimed damages for an explosion caused by the storage of oxygen cylinders. In the good character condition, the defendant had these to assist with the care of his mother, in the bad character condition, the cylinders were used to manufacture methamphetamine. Issues were foreseeability and causation. In the good defendant character condition, no judges found in favour of the plaintiff, but in the bad character condition 23.8% of judges did, a marginally significant finding. The issues supporting the outcomes were consistent with an

effect of character. For example: foreseeability 5.9% (good) v 23.8% (bad) and causation 35.3% (good) v 66.7% (bad), though these did not quite reach significance at the 0.05 level. Similarly to the previous experiment, only a single judge referred to character in their reasons.

### 3.2.3 The Legal Attitudinal Tradition

Research into legal decision making in the so-called 'attitudinal' tradition is often taken to be similar evidence of the impermissible influence of legally irrelevant attitudes on legal decision-making, but the research is not quite as compelling as that set out above. Characteristic of legal research in the attitudinal tradition is the robust finding that case outcomes appear to be positively associated with adjudicators' attitudes and outlooks. Thus, for example, more conservative minded judges in the US Supreme Court tend to vote for more conservative outcomes compared to more liberally minded judges hearing the same case (Epstein et al., 2013, pp. 77–78; Furgeson et al., 2008, p. 219; Pritchett, 1941, p. 892, 1948; Schubert, 1962, 1965; Segal & Cover, 1989; Segal & Spaeth, 1996b, 2002; Sheehan et al., 1992; Spaeth, 1961; Tate, 1981; Ulmer, 1960). Research in the attitudinal tradition makes clear that an adjudicator's outlook does not affect all of their decisions. Instead, for large areas of law, particularly criminal law, there is reasonable consensus such that adjudicators of all outlooks make similar decisions (Rachlinski et al., 2017, p. 2051; Sunstein et al., 2006, p. 61). It is only in a minority of cases raising issues of political or personal significance to the adjudicator that outcomes vary by the judge's outlook (Rachlinski et al., 2017, p. 2051; Rachlinski & Wistrich, 2017, pp. 208–209; Sisk & Heise, 2004, p. 746). Thus issues touching on politics (Maveety, 2003), religion, race (Broeder, 1959, p. 748; Cox & Miles, 2008, p. 1), gender (Peresie, 2004), and age are determined differently when these issues are also salient to the adjudicator (Rachlinski & Wistrich, 2017). The effect of a value also appears related to how strongly it is held by an adjudicator. Thus, while adjudicators will compromise on certain issues when sitting with other adjudicators holding different views (Cox & Miles, 2008, p. 1; Peresie, 2004, p. 1778), they are very much less willing to do so when these issues

touch on values that are particularly sacred to them, such as regarding abortion or the death penalty (Sunstein et al., 2006, p. 55).

Some care needs to be taken with using this as evidence of the effect of legally impermissible factors on legal decision-making because, as Spellman points out, while much of the attitudinal research is commendably based on the observation of real-life cases thus providing considerable external validity, this also means that the adjudicators' attitudes are not manipulated and therefore not random (Spellman, 2010, p. 162). Thus it is possible that a 'third variable' affects both attitudes and inferences. For example, evidence that adjudicators from gender and racial minorities often decide differently to other adjudicators when considering issues of gender and race is not necessarily evidence that they are influenced by legally impermissible factors. It could also be that their gender and racial background has made them aware of relevant matters that those from the majority are unaware. Yet the consistent directionality of the association, in that cases outcomes are almost invariably positively associated with outcomes that the adjudicator would prefer, does suggest that not all of the patterns will be explained by third variables.

Another reason for caution with research in the attitudinal tradition is that there are areas where the adjudicator is permitted a margin of discretion (Hart, 1958, 1961, pp. 12–13, 121–150; U. Moore & Hope, 1929, pp. 703–704; Schauer, 1988, p. 514). In such circumstances, it may be legitimate for the adjudicator to take some account of their personal view about the correct course of action. In the absence of quite a narrow forensic scrutiny to demonstrate that a particular issue was influenced by legally irrelevant factors, it can be difficult to conclude with high certainty that such an inference was impermissible.

Nonetheless, particular studies do seem to provide some support for the impermissible influence of attitudes on legal decision making. For example, in a jury context, Kahane et al asked participants to view a video of a political demonstration, and asked to determine issues such as whether those shown in the video had obstructed and threatened pedestrians (Kahan et al., 2012). Half of participants were told that the protest was outside an abortion clinic (a

more conservative cause) and half were told that the protesters were protesting against the US military's 'don't ask, don't tell' policy regarding sexuality (a more liberal cause). Participants political attitudes were also measured. Kahane et al found that participants of different attitudes assigned to the same condition disagreed sharply on their interpretation of key factual issues and also disagreed with participants with similar attitudes assigned to the opposite condition. Similarly, Sood and Darley found that participants punished a nudist protest differently, depending on whether the message that he was said to be conveying regarding abortion matched or conflicted with their own views (2012, p. 1339). At the very least, attitudinal research suggests that in appropriate cases, such as where there is a divergence of views as to the correct outcome, an adjudicator's outlook may provide them with a motive to influence case outcomes.

### 3.2.4 Disregard of Law and Instructions on the Law

Another aspect to the empirical picture is research suggesting that adjudicators will sometimes ignore the law or instructions on the law and instead do what they think is appropriate. Anecdotal evidence of decision makers ignoring the law was first available in relation to jury decision-making. In the 19th century, the famous judge and jurist OW Holmes observed that while the then law stated that an employer was only liable to an employee for negligence, if the case was allowed to go to the jury, the jury would generally find for the employee, even where there was no negligence. Holmes' explanation was that the general intuition on the part of the layman was that employers should insure the safety of those they employ (Holmes, 1897, p. 466). Myers similarly found evidence of real-life juries taking collateral information into account to such an extent that they seemed reluctant to convict on the basis of distinctions not recognised by the law such as the status of the accused and victim (Myers, 1978, p. 795). Where the law treats a prior criminal record as relevant only to a defendant's credibility rather than propensity to commit crime, Wissler and Saks found that mock jurors found that a prior record nonetheless affected propensity rather than credibility (Wissler & Saks, 1985, pp. 43–44). Eisenberg and Hans similarly found from a

survey of court proceedings that in weak cases, disclosure of a prior criminal record was linked to higher conviction rates, suggesting that jurors were nonetheless using a criminal record to assess propensity (T. Eisenberg & Hans, 2008, p. 1353). Relatedly, Kassin & Sommers found that mock jurors did not always ignore evidence ruled inadmissible, despite being instructed by the putative judge to do so (Kassin & Sommers, 1997, p. 1046). Of particular relevance is that the jurors took the basis of the ruling into account in a principled way when deciding to rely on it or not: thus evidence ruled inadmissible for being illegally obtained led to more convictions than evidence ruled inadmissible due to unreliability. While it might be tempting to put such phenomena down to the lack of experience and expertise on the part of jurors (Wistrich et al., 2004, pp. 1251–1252, 2015, p. 900), research in fact suggests that judges have a similar tendency to ignore the law. Thus Landsman & Rakos found that both juries and judges were inappropriately influenced by inadmissible material (Landsman & Rakos, 1994, pp. 122–123). Wistrich et al equally found that judges had difficulty disregarding demands disclosed during settlement discussions, privileged conversations, prior sexual history, prior criminal convictions, or information subject to undertakings it not be used (Wistrich et al., 2004). One exception is that judges did seem to be able to resist information that they were legally prohibited from considering where this directly implicated constitutional rights. However, these types of cases were ones that were more likely to be appealed to a higher court if erroneous whereas the other scenarios were much less likely to be subject to supervision by another court. This meant that these types of decisions would have been much more closely scrutinised, which may have explained the difference.

### 3.2.5 Influence Proportionate to Ambiguity

Ambiguity of the decision seems to be an essential ingredient for extralegal information to have an influence. Adjudicators seem to be influenced by proscribed information where the nature of the decision is ambiguous, which correspondingly makes it likely that the influence of the proscribed information is difficult for observers to discern

(Braman, 2006, 2009; Braman & Nelson, 2007; Wistrich et al., 2015, p. 900). As noted above, this often makes it almost impossible to discern the influence on an individual judge in an individual case. The most that can often be said is that statistically it is likely that the information had an effect on a proportion of judges, but it cannot be said for certain which individuals (Wistrich et al., 2015, p. 904). This link with ambiguity again echoes findings in the related psychological field of motivated reasoning. Thus Kunda writes that motivated reasoning is constrained by what would persuade a dispassionate observer (Kunda, 1990, pp. 482–483) and Klein, & Kunda point out that people's desires affect their conclusions only if they can construct rational justifications for them (Ditto et al., 2009, p. 312; Hsee, 1996, p. 122; W. M. Klein & Kunda, 1992, p. 146). This seems to be what we see in the legal context, with Wistrich concluding that judges are influenced by extralegal factors only where the law is unclear (Katz & Spohn, 1995, p. 178; O'Neill, 1981, p. 626; Posner, 2008, pp. 132, 137; Wistrich et al., 2015, p. 900). Thus, for example, Braman found that participants' preferences regarding abortion had more of an influence on the legally unrelated issue of standing to sue where there was less legal precedent on the issue (Braman, 2006, pp. 319–320). As many commentators also point out, the related *ex post facto* obligation to provide reasons discussed at Section 2.9 may act as a further constraint on an adjudicator because reason giving tends to make a decision more transparent and less ambiguous (M. A. Eisenberg, 1978, p. 412; Fuller, 1978, p. 388; Knight, 2009, p. 1550; Posner, 2008, pp. 110–111; D. L. Shapiro, 1986, p. 737). Consistent with such a view, Liu found that judges who provided written reasons for their decisions were less influenced by extralegal information (Liu, 2018, p. 83). Finally, evidence shows that if the ambiguity is removed by presenting counterfactual cases side-by-side, the effect of extralegal information practically disappears (Cushman et al., 2006; Nadler, 2012, p. 26; Sood & Darley, 2012, pp. 1343–1344).

### 3.2.6 Changes in Representation of Collateral Information

Discerning the influence of extralegal information on legal decision making is made more difficult because adjudicators seem to present their thought processes as more coherent

than they actually are. When fact-finding or decision-making, adjudicators make inferences on the basis of more universal generalisations that antecede the facts of the case before them. For instance, in a homicide case, an adjudicator asked to infer whether carrying a knife is evidence of premeditation might conclude that it is or is not evidence of premeditation based on the generalisation either that it is abnormal to carry a knife or normal to do so (N. Pennington & Hastie, 1991, p. 556). When an adjudicator gives reasons for their decision, they are expected to explain how they reached their conclusions, including some detail of the generalisations that they relied upon. Ordinarily, disclosing these generalisations provides some insight into a decision, including sometimes whether extralegal information was impermissibly taken into account (M. A. Eisenberg, 1978, p. 412; Fuller, 1978, p. 388; Knight, 2009, p. 1550; Posner, 2008, pp. 110–111; D. L. Shapiro, 1986, p. 737). For example, if an adjudicator says in their reasons that they believe it is abnormal to carry a knife, but nonetheless concludes that a killing was not premeditated, an observer might find this suspicious and perhaps begin to explore more closely whether impermissible factors influenced the decision (Engel, 2006, p. 250; Sperber, 2001, pp. 409–410; Thompson, 1985; Walton, 2005, p. 48). But research, particularly by Dan Simon and his collaborators, suggests that adjudicators may also manipulate the generalisations they give in their reasons to fit with their decisions. Simon et al showed this by eliciting views from participants on the types of generalisations that would be relied upon in a later legal case before the participants saw the actual case. They then compared the participants' reports of the same generalisations after they had seen the case. Though the generalisations, being more universal, would be expected to be stable, Simon et al actually found that participants' views on the generalisations seemingly changed so as to fit with their decision on the case. Thus participants assessments of the analogy of whether the internet was more akin to telephony or a newspaper were quite equivocal in the abstract before they had seen the case. But once they had seen the case, their assessments of the analogy invariably changed to be much more polarised and, crucially, to agree with their final verdict (Glöckner & Engel, 2013, p. 245; Holyoak & Simon, 1999; D. Simon et al., 2001; D. Simon, Snow, et al., 2004). Simon argues that the same pattern can be discerned in real life cases (D. Simon, 1998, pp. 19–20, 83). This also seems to be a pattern that is also observed in the psychological context of motivated reasoning (Kunda, 1990, p.

483; Wicklund & Brehm, 1976). The overall effect of this associated phenomenon is that it results in an 'impoverished discourse' (D. Simon, 1998, p. 121) that 'deprive[s readers] of any possibility of distinguishing between good and bad arguments' (D. Simon, 1998, p. 130). Interestingly, this apparently drastic change in outlook also seems to be temporary: when the same participants views are elicited some weeks after the initial exposure to the extralegal information, their views appear to revert back to their pre-experimental exposure values (D. Simon & Spiller, 2016, p. 1588).

### 3.2.7 Apparent Lack of Insight

Intriguingly, adjudicators whose outlooks on these general points of principle seem to change so drastically between the point when articulate their views before being exposed to the extralegal information and the point after they have seen the extralegal information appear to have little conscious insight that their views have apparently changed so drastically (but see Glöckner & Engel, 2013, p. 245; Holyoak & Simon, 1999, p. 18; Posner, 2008, pp. 69–70; D. Simon, 1998, p. 61, 2004, p. 533). Researchers are uncertain whether decision makers are consciously aware that they have been influenced by impermissible information (Braman & Nelson, 2007, p. 954; Liu & Li, 2019, p. 657; D. Simon, 2010, p. 142; Spellman, 2010, p. 162; Wistrich et al., 2015, p. 899), though given such a question relates to an internal state of mind that a decision maker might be motivated not to reveal, such a question might be challenging to investigate empirically.

### 3.2.8 Early Impact Even When Information is Viewed Passively

Such legally irrelevant information seems to have an impact very early in the adjudicatory process, long before the end point where an adjudicator is expected to give reasons, and even in situations where participants review a case in a capacity other than as an

adjudicator. This seems contrary to the expectations of some theorists. For example, to the extent that adjudicators might be taking into account impermissible considerations, there are some parallels with Festinger's 'cognitive dissonance theory' (Festinger, 1962). Festinger theorised that in making some decisions, individuals processed information and assessed alternatives purely objectively. It was only at the point when they were committed to a decision, and presented with information that gave rise to 'cognitive dissonance' that they allowed their decision to be influenced by impermissible considerations (Festinger, 1962; D. Simon, 1998, p. 53). Contrary to Fesinger's view, Simon et al's research indicates that the irrelevant material seems to be influencing the adjudicatory process at a very early point. For example, in adjudicatory experiments, Holyoak and Simon included a condition where some participants were provided with partial evidence, told that further evidence was due to be received, and asked to give a non-binding 'preliminary leaning' of the case (Holyoak & Simon, 1999). Notwithstanding that the evidence was incomplete, and it was clear that a final assessment was some way off, participants assessments were nonetheless still influenced by the legally irrelevant material. Follow up experiments by Holyoak et al replicated this effect and also confirmed that collateral factors influenced participants at an early stage even where it was presented as a memory or comprehension test rather than any form of legal assessment (Furgeson et al., 2008, pp. 224–225; D. Simon et al., 2001, p. 1250).

### 3.2.9 Moderating Factors

In addition to the factors outlined above that influence whether extralegal factors affect decision making, there are also a number of factors that seem to moderate the effects. As noted earlier, decisions that are expected to be scrutinised more closely reduce the impact of extralegal information (Kunda, 1990, p. 481; Tetlock, 1983; Wistrich et al., 2004, p. 1324). Where decisions are made in panels, the existence of even a single adjudicator of the opposite persuasion seems to moderate the position of the majority (Cox & Miles, 2008, p. 1; Liu, 2018, pp. 85–86; Peresie, 2004, p. 1778; Posner, 2008, p. 31; Rachlinski & Wistrich, 2017, p. 209; Sunstein et al., 2006, pp. 54, 64). Pre-existing legal precedent seems to reduce the effect

(Braman, 2006, pp. 319–320; Johnson, 1987, pp. 338–339), possibly because it reduces the range of reasonable decisions that an adjudicator could come to, thereby attenuating the available ambiguity (Schauer, 1988, p. 510). Likewise, the obligation to provide reasons to explain a decision has long been recognised as a means of making the underlying decision more transparent to observers (M. A. Eisenberg, 1978, p. 412; Fuller, 1978, p. 388; D. L. Shapiro, 1986, p. 737). Correspondingly, requiring an adjudicator to give reasons tends to lessen the impact of character information (Tetlock, 1983), save where it is not the adjudicator who is giving reasons but somebody to whom they delegate the task (Liu, 2018, p. 83). Finally, incentives to be accurate sometimes moderate the impact of collateral information , but not all the time (Furgeson et al., 2008, p. 219).

## 3.3 THE ISSUES

The ultimate issue between the prosecution and the defence is how to explain these empirical patterns. The case for the prosecution is that these patterns are a result of shortcomings in human and judicial rationality (Holyoak & Simon, 1999, p. 23; D. Simon, 1998, 2004; D. Simon, Snow, et al., 2004, p. 822; Zamir et al., 2014, p. 675). Essentially, allege the prosecution, when adjudicators' decisions are influenced by extralegal factors, this is a failure of rationality, sometimes said to be caused by the complexity of the task or the limited cognitive capacity of the adjudicator (D. Simon, 1998, p. 121). Often this argument is put in terms of 'dual-process' theories of cognition as set out at Section 2.8.2 above. Such arguments suggest that adjudicators would make rational decisions if they used characteristically 'deliberative, rule-governed, effortful, and slow' 'System 2' reasoning which is, but due to the difficulty of the challenge and their cognitive limitations, they in fact slip into using characteristically 'spontaneous, intuitive, effortless, and fast' 'System 1' reasoning (Liu, 2018, pp. 88–89; Posner, 2008, pp. 107–108; Rachlinski & Wistrich, 2017, p. 223; Spottswood, 2013; Wistrich et al., 2015, pp. 863–864).

By contrast, the case for the defence is that these patterns may be evidence of a

characteristically rational response to the problems faced by an adjudicator. And far from being evidence of the limitations of judicial rationality, they might be the hallmarks of a sophisticated rationality by which some adjudicators undertake activities that go well beyond the recognised fact-finding and decision-making responsibilities that they are tasked with. In summary, there are situations where a standard application of the law to the facts would lead to a case outcome that the adjudicator would not favour. This will sometimes trigger an assessment of the scope that that adjudicator has to secure an alternative outcome that they would prefer. Alternative outcomes necessarily entail presenting impermissible inference processes as permissible ones. This requires the adjudicator to consider the prospects of other parties identifying such behaviour. Such analysis encompasses both the inherent ambiguity of the context, as well as the measures that the adjudicator could take to increase this ambiguity, such as misrepresenting the generalisations that are driving their decisions. Nonetheless, to say that the behaviour of adjudicators in such circumstances is rational does not mean that it is objectively desirable or should be encouraged. Instead, seeing the behaviour as rational provides a theory for predicting the circumstances in which it will occur, and understanding the means that might be available to influence it. Finally, suggesting that these empirical findings are not necessarily evidence of irrationality does not absolve adjudicators of accusations of irrationality in other circumstances.

## 3.4 THE DEFENCE CASE

Defending judicial reasoning against assertions of irrationality will take the form of propounding a positive alternative case. Rather than bare denial, or putting the prosecution to proof, the defence will outline an alternative explanation for the striking empirical patterns that we see. In doing so, we will examine three common pillars of a criminal investigation: means, motive, and opportunity. The alternative case that will be put forward is that adjudicators do take into account impermissible collateral information where it enables them to arrive at outcomes that they prefer and that they also take steps to conceal that they have done this. But before we start this task, some preliminary comments on the appropriate way

to assess rationality are needed.

### 3.4.1 The Correct Reference Environment for Rationality

The first preliminary point to note is that when trying to understand the behaviour of an adjudicator, we should try to understand that behaviour in the context of real-life adjudication, not the artificial experimental environment that psychologists create to try to glean more information about that behaviour (H. A. Simon, 1956, pp. 129–130; Todd & Gigerenzer, 1999, p. 13). A real-life environment is often very different from an experimental environment. Psychologists try to create experimental environments that are more transparent than the real world with techniques such as taking larger samples to detect statistical patterns and introducing counterfactuals and randomisation. Particularly where the rationality of a behaviour relies on being covert rather than overt, what may be rational in the oblique real world environment may not be rational in the transparent experimental environment.

The second preliminary point worth noting is that we should also be careful not to uncritically introduce the law as a standard to assess the rationality of judicial behaviour. While the law often reflects widespread values in the community, it does not always do so. There are situations where the law is contested or imperfect. These situations might cause an adjudicator to seek an outcome that is not as prescribed by law. To seek a different end to the law, and to find means to do so, is not necessarily irrational. Rather, only once we feel we have a good descriptive understanding of psychological behaviour such as adjudication should we introduce a normative standard to assess that behaviour, whether it be the law or any other system of values.

### 3.4.2 Motive

Means, motive, and opportunity is a triumvirate that investigators or prosecutors use to understand or communicate why somebody an accused might have committed a crime (Fenton & Neil, 2013, p. 419). We will use these three concepts to look at how the effect of extralegal information on case outcomes might be explained as a rational response by the adjudicator.

Taking the investigatory concepts somewhat out of turn, logic and the evidence set out above does suggest that judges are motivated to seek their preferred outcomes. At the very minimum, outcomes are reliably correlated with adjudicatory outlook. Thus conservative judges are associated with conservative outcomes, and liberal judges with liberal outcomes; minority judges favour minority outcomes; and most adjudicators favour good characters over bad characters. This association is particularly clear in the context of the attitudinal tradition of legal research even if, for reasons explained previously, it is difficult to say that they always take these values into account impermissibly. We also see that the strength of the motivation appears related to how strongly the values are held. Thus adjudicators sitting with other adjudicators are very reluctant to compromise the outcome where it relates to a particularly sacred value, such as mortal values about the status of the embryo or to the death penalty (Sunstein et al., 2006, p. 55). Given the existence of a motivation in circumstances where there is a lawful opportunity for an adjudicator to favour outcomes that matter to them, it is not a great leap to suggest that there remains a motivation to favour these outcomes where it would be strictly unlawful to do so.

### 3.4.3 Opportunity

A second consideration is opportunity. Opportunities to take extralegal information into account when determining outcomes are obviously limited when to do so would be unlawful. Nonetheless, they will still sometimes exist. The empirical evidence confirms that

the use of extralegal information is often proportionate to the size of the window of opportunity. Thus we only see the influence of extralegal information where there is some ambiguity in the nature of the issue that the adjudicator has to determine (Braman, 2006, 2009; Braman & Nelson, 2007; Wistrich et al., 2015, p. 900). Where there is ambiguity in the factual inference, drawing of analogy, law, or decision making, this leaves the door open for an adjudicator to determine the issue in a way that would favour the outcome that they are more sympathetic to. By contrast, where there is no ambiguity in the issue for the adjudicator to determine, favouring an unnatural outcome would immediately look suspicious to observers. In accordance with this, extralegal information has been shown to practically no effect where the outcome is obvious.

Precedent and the obligation to provide reasons also have the effect of closing down the opportunities for an adjudicator to be influenced by extralegal information. Where there is much precedent on an issue, there is less discretion available to the adjudicator (Ho, 2008, p. 35; Schauer, 1988, p. 521) and the influence of extralegal information is correspondingly more limited. Similarly, a requirement to provide reasons for a decision after the fact makes a decision more transparent. This, in turn, limits the options for a judge to be influenced by extralegal information. Correspondingly, a requirement to give reasons tends to diminish the effect of extralegal information. Where ambiguity caused by the absence of counterfactuals is resolved by presenting cases side-by-side, the effect disappears (Cushman et al., 2006; Nadler, 2012, p. 26; Sood & Darley, 2012, pp. 1343–1344).

The level of scrutiny that a decision will be subjected to also has the effect of opening up or closing down the window of opportunity that might be available to an adjudicator. Correspondingly we see that extralegal information has more of an impact where scrutiny levels are low. Thus, as Wistrich et al show, it is in the cases raising constitutional rights that tend to be appealed and thereby subjected to higher levels of scrutiny that extralegal information has less of an impact than in cases unlikely to be scrutinised so carefully (Wistrich et al., 2004).

### 3.4.4 Means

In terms of the means by which an adjudicator may take advantage of the opportunity to further their motives, there are two symbiotic aspects to this. First is the primary aspect of taking into account extralegal information in their decision making. But blindly taking into account extralegal information would not be terribly effective because it leaves the door open to the strategy being unmasked. Thus a secondary aspect to the strategy is needed. This is to manipulate and impoverish the information available to observers so that it is more difficult to reliably discern the primary influence of extralegal information.

The primary aspect is most easily understood in the context of a legal decision where the decision maker is not obliged to give reasons (Schauer, 1995, p. 634). In contested first-instance cases, this means that the adjudicator may follow quite a complicated series of inferences from the evidence to the facts and then from the facts to the decision, at each stage taking into account the applicable law. But this means that the only information available to an observer is the evidence and final decision. With such limited information, it becomes very difficult for an observer to infer whether the adjudicator has impermissibly taken extralegal information into account. The adjudicator may thereby be granted significant autonomy to select the outcome they prefer given the extralegal information. Instead of choosing the outcome that they would tend to select absent the extralegal information, the adjudicator can instead choose another option. Consistent with this, we see that where there is no obligation to give reasons, there tends to be a greater effect of extralegal information (Liu, 2018, p. 83; Tetlock, 1983). While not all adjudicators seem to be so influenced, a significant proportion often are. It may therefore be unsurprising that the obligation to give reasons is an immutable component of the right to fair trial, recognised in most jurisdictions (Feteris, 2017, pp. 18–19; Hirsch, 2003, p. 618; Knight, 2009, pp. 1543–1544; Posner, 2008, pp. 110–111).

The associated secondary aspect comes into play when the adjudicator is faced with more challenging environments, in particular where there is an obligation to disclose reasons

for the decision, either in oral or written reasons to the public, or in reasons to other members of the bench. As we have observed, giving reasons makes the adjudicator's inference process more transparent (Engel, 2006, p. 269; Knight, 2009, pp. 1543–1544; Posner, 2008, pp. 110–111; D. Simon, 1998, pp. 12–13) because each link in the chain of inferences can be checked one at a time. A fallacious inference is associated with a mistake, carelessness, or an error of law such as the influence of extralegal information(Engel, 2006, p. 250; Thompson, 1985), and can be grounds to overturn the decision (Posner, 2008, p. 81; Sperber, 2001, pp. 409–410; Walton, 2005, p. 48). To counter this, the adjudicator can make it more difficult for an observer to identify inconsistencies by manipulating and impoverishing the information available in their reasons (D. Simon, 1998, pp. 121, 130). In particular, they may misrepresent the generalisations that they rely on to reach the decision in order to ensure that the generalisations are consistent with the decision and the law, if not with their actual inferences. Thus, for example, where an adjudicator favours a party due to the extralegal factor of their good character rather than a strong feeling on the delicately balanced issue of whether the internet is more like a telephone system or a newspaper, they might conceal this behaviour to some degree by saying in their reasons that they in fact feel strongly that the internet is more like a telephone system (Holyoak & Simon, 1999, p. 9). In an isolated case, such behaviour is not discernible, but with a randomised larger sample it can be inferred statistically. Finally, adjudicators can deny that they have changed their views in the light of extralegal information, acting as though they had always held such views (but see Glöckner & Engel, 2013, p. 245; Holyoak & Simon, 1999, p. 18; D. Simon, 1998, p. 61, 2004, p. 533). This further reinforces the difficulty for an observer to discern the change of views. Thus, such behaviour is not an irrational inference that the internet is more like a telephone because the party is of good character, but the rational obscuring of the impermissible inference that the party should win because he is of good character. The resulting 'impoverished discourse' may therefore be a feature, not a flaw.

## 3.5 THE PROSECUTION CASE

### 3.5.1 Directionality of Outcomes

Turning to the prosecution, the story is that the distinct empirical patterns that we see are caused by failures of rationality due to the challenge of the cognitive complexity of the adjudicatory task compared to the cognitive complexity of the adjudicators. But if this were the case, we ought to expect case outcomes to be in random, rather than oriented, directions. Just as each unhappy family is unhappy in its own way, each failure of rationality might be expected to favour one side just as much as another. Yet the evidence does not seem to show this. Rather, case outcomes invariably seem to favour the outcome or party that the decision maker favours. This does seem to point to the influence of motive, whether conscious or unconscious, rather than coincidence.

### 3.5.2 Case Simplicity

Similarly, if the prosecution explanation is that these effects are partially because of the complexity of the task, it might be reasonable to infer that we should see more significant effects of extralegal information in complicated cases compared to simple cases. But this does not seem to be the situation. As the empirical evidence shows, these effects seem to occur across the board and even in relatively simple cases. It seems somewhat unlikely that the rationality of adjudicators is consistently failing even in quite basic cases. Instead, the touchstone seems to be the combination of ambiguity in the decision to be made, combined with a motivation triggered by extralegal information. Overall, this seems less consistent with the prosecution's irrationality hypothesis and more consistent with the defence case previously outlined.

### 3.5.3 Point where Manipulation Arises

One of the strongest arguments relied on by the prosecution in favour of the irrationality hypothesis is that the patterns we observe arise at interim stages, well before the evidence is complete. This is taken to be incompatible with the defence hypothesis that the patterns are caused by adjudicators manipulating secondary information to obscure the influence of extralegal information. However, the theory that this empirical pattern is taken to knock down is related to our hypothesis, but not identical with it. Festinger believed that the manipulation of secondary information about decision-makers' thought processes only happened once the decision maker was 'committed to a decision' (Festinger, 1962; D. Simon, 1998, p. 53). Dan Simon translates Festinger's position to the legal adjudicatory context, and takes it as implying that any manipulation of secondary information would only happen once all the available evidence was complete (Holyoak & Simon, 1999, p. 4). Because the assessments of the evidence seem to be influenced by extralegal information at a very early stage and before all the evidence is complete, this is obviously incompatible with Festinger's theory (Holyoak & Simon, 1999, p. 21).

However, while it might be natural to assume that manipulation of information would take place only once all the evidence was available, careful analysis suggests that such an assumption might not be safe. Instead, a better strategy for an adjudicator could be to take account of key extralegal information as soon as it is presented in order to engage in the secondary strategy of managing and manipulating information about their thinking process immediately. Admittedly, this would be more cognitively challenging for it would require the adjudicator to repeatedly review how they would manipulate information about their thought processes in the light of each new piece of evidence. But by contrast, it would seem to offer notable advantages. In particular, if sitting as a panel, it would allow the adjudicator to be ready to influence other adjudicators with a coherent but biased account from the outset. Furthermore, given that an adjudicator might be asked at any interim stage for an account of their thinking, it would seem essential that an interim view is coherent with a final view. The discrepancy between giving an account unbiased by extralegal information at an interim

stage, but a final biased account at the end would risk their manipulation being readily revealed. A safer strategy would be to refer consistently to a biased account. Finally, it ought also to be recognised that it may be difficult to identify the point at which the evidence is definitively complete. Expected evidence may not materialise, and new evidence may be presented unexpectedly. In such an environment, a better strategy would also seem to be to refer to a biased model at all times. This account would explain the empirical observations compatibly with the theory set out here.

Relatedly, the prosecution also rely on the fact that we see the same empirical patterns even when participants are presented with the same information outside an adjudicatory context, for example when the cases are presented as memory or comprehension tests (Furgeson et al., 2008, pp. 224–225; Holyoak & Simon, 1999; D. Simon et al., 2001, p. 1250). If the defence say that the behaviour is aimed at promoting outcomes that the participant favours, it seems somewhat incompatible with the theory for the participant to behave in this way when there is no apparent means of influencing any outcome. In defence response, it is appropriate to note that adjudication is not the only arena where a person can influence their social environment. Receiving and passing on information and opinions in a social context also amounts to a means of influencing the environment (Sperber et al., 2010). Gossip that exaggerates misbehaviour may be an effective means of punishing known bad characters. Thus, though the experiments presented may have been described as memory or comprehension tests, the substance of the information must have inevitably triggered senses of empathy or antipathy. The descriptions may well have been insufficient to overcome important intuitions to manage information, and thus triggered the same patterns as when they were described as legal cases.

### 3.5.4 Coherence Absent Instrumental Goal

A final pattern relied upon by the prosecution is that adjudicators sometimes seem to manipulate secondary information about their decision making process even where there is

apparently no relevant extralegal information or side that would attract sympathy or antipathy. Thus there are some circumstances where adjudicators seem to prefer to exaggerate the coherence of their reasoning process for apparently intrinsic reasons. Thus, adjudicators seem to present the secondary 'impoverished discourse' even when there is no primary behaviour to be concealed. Simon discusses the stark contrast between the numerous contested issues in a case during the argumentation stage compared to the apparently incontrovertible yet opposing opinions written by judges to explain their decisions where judges on one side settle all arguments one way and the judges on the other side settle all the arguments the other way (D. Simon, 1998, pp. 19–20). Note that Simon refers to arguments that are logically and legally unrelated, so there ought to be no reason for these arguments to be aligned. Simon observes that such a diametrically opposite alignment of arguments is 'plainly implausible' (D. Simon, 2010, p. 139). This inherent preference for coherence seems more challenging to account for. One possible theory is that the process of making it difficult for an observer to scrutinise an adjudicator's reasoning process is a by-product of other circumstances when there is behaviour to conceal. Thus, because inconsistencies in an adjudicator's reasons may be taken as evidence of bias, carelessness, or ignorance (Engel, 2006, p. 250) and being overturned is therefore a blow to their reputation (Posner, 2008, p. 81), adjudicators might therefore engage in this secondary process of manipulating information to be more coherent as a matter of course to avoid the appearance of bias, carelessness, or ignorance.

## 3.6 CONCLUSION

The purpose of this section has been to suggest that legal adjudicators may not always be guilty of the crime of behaving irrationally. To try to raise a reasonable doubt, we have not just challenged the irrationality hypothesis, but suggested a possible alternative case. This alternative case is that the empirical patterns that we see may be a hallmark of rationality. While adjudicators may be irrational for some of the time, there is perhaps a logic behind some of the empirical patterns that we see, a logic which would tend to realise adjudicator's

goals in the real-world adjudicatory environment, if not in a psychological laboratory. It seems possible that a sizeable proportion of adjudicators may realise their goals by taking advantage of ambiguity in the decision-making environment and by manipulating information about the decision-making environment to make their behaviour less obvious to observers. Where there is extralegal information available that the adjudicator is sympathetic to, and the nature of the task is sufficiently ambiguous for that adjudicator to take that extralegal information into account by arriving at a decision that they prefer, a proportion of adjudicators seems to do so. These processes seem to take place as soon as the extralegal information is presented and seem sufficiently ingrained that they will be triggered by presentation of conflicts, even if these disputes are described as memory or attention tests.

In trying to exonerate adjudicators of the charge of irrationality, it might be thought that adjudicators have instead been implicated in alternative charges, such as unlawful behaviour or misrepresentation. Superficially, there appears to be merit in such an argument, but considering the law more widely, a more nuanced position might be reasonable. True, such behaviour would be less transparent and less accountable that might be desirable, and may sometimes lead to arbitrary outcomes. But it is also easy to see circumstances where taking extralegal information into account might be merited. Formal law is not always perfect: indiscriminate laws rarely foresee every eventuality that may come before a court, and in some legal systems the laws may be deliberately designed to be oppressive. In those circumstances, an adjudicator taking advantage of ambiguity and extralegal information may be able to inject a little humanity into proceedings. Overall, a better approach might be to improve our psychological model of adjudication, recognising that written laws may not always be applied in the ways we assume.

A more accurate psychological model of adjudication would also lead to policy prescriptions that are more likely to be effective. The prosecution's model of adjudication based on the irrationality hypothesis would suggest that we should be training and supporting adjudicators to make better decisions. By contrast, the alternative defence model suggested here would suggest that such measures would be unlikely to be effective, and that it would be

more effective, where possible, to either reduce the ambiguity in the legal issues being determined, or to insulate adjudicators from finding out about the types of extralegal information known to influence decisions. In the following section, we will seek to test this alternative view of adjudication experimentally.

# 4. THE EFFECT OF LEGALLY IRRELEVANT INFORMATION ON OUTCOMES

## 4.1 INTRODUCTION

At this point, we turn from a review of the existing literature and extension of the existing theory to empirically testing the more fleshed out theory that we have developed. In particular, the following section focusses on the influence of what we defined in s.2.5 above as 'extralegal' information, information that does not appear to further the general values of a legal system. Our interest is in exploring whether the empirical evidence we can reveal is more consistent with a picture of adjudicators behaving irrationality due to cognitive limitations or task complexity, or whether it is more consistent with a more rational picture, whereby adjudicators favour a party they are more sympathetic to, even if this is on the basis of extralegal information, and behave in ways that make this difficult to discern in the real-world environment.

## 4.2 STUDY 1

In Study 1, we explored the effect of a defendant's character on the determination of a single factual issue that from a legal perspective had no plausible link with character, a so-called 'extralegal' effect (Posner, 2008, p. 42; Rachlinski & Wistrich, 2017, p. 205; Rowland & Carp, 1996, p. 136; Segal, 1984, pp. 899–900; Wistrich et al., 2015, p. 900). One of the goals of this was to isolate the findings of previous research that suggested that extralegal considerations such as character may nonetheless influence the determination of ambiguous issues (Holyoak & Simon, 1999, p. 21; 2019, p. 637; Wistrich et al., 2015). In addition, given the argument from some theorists that such phenomena amount to irrationality, either because the scale of the cognitive challenge exceeds the cognitive capacities of the decision maker (Holyoak & Simon, 1999, p. 23; D. Simon, 1998, 2004; D. Simon, Snow, et al., 2004, p. 822;

Zamir et al., 2014, p. 675), or because decision makers are using heuristic shortcuts rather than complete analysis of the legal case (Liu, 2018, pp. 88–89; Posner, 2008, pp. 107–108; Rachlinski & Wistrich, 2017, p. 223; Spottswood, 2013; Wistrich et al., 2015, pp. 863–864) we deliberately chose a very simple paradigm with very little complexity. This paradigm required the decision maker to assess the single issue of whether the location of the alleged wrongdoing was public or private from an ambiguous photo, an issue which correspondingly determined whether this amounted to a crime or not.

Given that our survey participants were lay members of the public, we chose an adjudicatory function often discharged by laypeople, namely the magistrates' courts. One difference between laypeople generally and lay magistrates is that the latter may be familiar with legal precedent and this could affect responses. To attempt to minimise this influence, we chose a rarely prosecuted offence with little or no material precedent, that of failing to clear up dog fouling contrary to a public spaces protection order imposed pursuant to sections 59-61 of the Anti-Social Behaviour, Crime and Policing Act 2014. As the name of the offence indicates, dog fouling cannot be criminalised unless it occurs in a 'public place'. The issue of whether the location of the crime was in a public place was chosen as the single issue to be determined because it bears no legal relationship to the character of the accused. In addition, the issue was found to be sufficiently ambiguous (on the basis of a small pilot survey of 10 individuals) that reasonable people could settle that issue either way. All other issues were presented as agreed between the parties, and the accused gave no evidence.

In accordance with our theory that this phenomena is evidence of rational behaviour in the usual adjudicatory context in that it allows a decision maker to favour a party that they have sympathy with without this behaviour being discerned, we hypothesised that despite character being on its face legally irrelevant to whether a location was public or private, a significant proportion of participants would nonetheless take character into account when deciding the case, and would correspondingly determine the issue and give reasons consistent with that decision.

### 4.2.1 Method

#### 4.2.1.1 Participants

One hundred participants recruited using the online survey platform Prolific completed the survey (64 females, 36 males; aged 18 to 60, M age 33.7, SD 10.0; 27% were students; 50% in full time employment, 29% in part time employment, 21% unemployed or other) and were paid £0.50 for their time. The sample size chosen had 80% power to detect an effect size with OR = 3.45 in a two-tailed Fisher's Exact test. To reflect the qualification requirements of a lay magistrate, participants were selected on the basis of British nationality and residency in England and Wales. Because of an unidentified technical issue, one participant was able to complete the survey twice. This participant's second participation was therefore excluded from the analysis. The study was conducted in accordance with approval obtained from UCL's Research Ethics Committee (CPB/2014/006). Informed consent was obtained from each individual in advance of participation by providing them with study information and a consent form to agree to.

#### 4.2.1.2 Design

We used a between participant design with one independent variable (character) with two levels (good and bad) in which all participants viewed a single set of materials where the putative accused was either of good character or bad character.

#### 4.2.1.3 Materials

The materials amounted to a realistic transcript of a trial in a magistrates' court in

England and Wales, presented using the Qualtrics online survey platform. Participants read the charge of allowing a dog to foul in a public place and failing to remove the waste accompanied by particulars of the offence. They then read a transcript of the evidence which consisted of the prosecution evidence given by a witness, an enforcement officer who had witnessed the events that were the subject of the charge. This prosecution witness was first examined-in-chief by the prosecution and explained what he witnessed. The witness also exhibited a photograph of the place where the events were said to have taken place. This photograph showed a somewhat ambiguous area in the corner of a residential area, behind the obvious pavement and unenclosed, but with various notices attached to the wall behind the area including ones that read 'no parking' and 'private'. The witness marked on the photo where the alleged offence was said to have taken place. This was then followed by a transcript of the cross-examination where the enforcement officer was cross-examined by the defence. Participants were informed that the defendant did not give evidence.

Participants then read transcripts of the closing arguments by the prosecution and defence in which both sides set out why they said that the issue of whether the area in question was public or not should be determined in their favour. Finally, participants read some advice on the applicable law said to have been provided to the legal clerk to the justices. This underlined that it was agreed that the accused's dog fouled, that she did not remove the waste, so that the single issue was whether or not the participants were satisfied to the criminal standard that the place in question was a public place.

At the conclusion of the case, participants were asked for their view on the issue, their verdict, and an explanation of their verdict. They were able to refer back to the materials when considering their decision.

The character manipulation amounted to a single off-hand response to a question posed by the defence. In the good character condition, the prosecution witness mentioned that the accused was elderly, did not have her glasses, and was very apologetic to the witness. In the bad character condition, the prosecution witness mentioned that the accused told the

witness to go away, used offensive language, and said he should pick up the waste himself.

#### 4.2.1.4 Measures

Participants were required to indicate a binary response to the issue of whether they were satisfied that the area was public or not public. They were then required to indicate a binary response to whether they found the defendant guilty or not guilty. The order of presentation of each of these responses was randomised by the survey platform. For external validity as well as to gain a qualitative insight into participants' thinking, participants were asked to give reasons for their verdict, as a lay magistrate would be required to do, using an open-ended text box.

Participants were also asked what the issue in the case was, and asked to choose between three options, which were also randomised by the survey platform. For those participants that gave an inconsistent answer, such as concluding that the area was public but finding the accused not guilty, or concluding that the area was not public but finding the accused guilty, an extra question was posed asking them to explain the inconsistency using an open-ended text box.

#### 4.2.1.5 Procedure

As previously noted, participants were recruited online and participated in the survey in a place of their choosing, using their own device. On referral from the Prolific platform, participants were first provided with the study information form. They were then asked to complete the consent form comprising a number of statements to which an affirmative answer was required in order to participate in the survey. An anonymous user identification was collected to enable subsequent matching of demographic data without compromising the

participants' anonymity.

Participants were randomly assigned to one of the two conditions by the Qualtrics survey platform and thereafter viewed a single version of the two versions of the case described above, each of which was identical other than the text of the character manipulation presented during the cross-examination transcript.

After reviewing the transcript, participants were asked to complete the measures described in the previous section and the additional measures, if applicable.

After completing the survey, participants were thanked for their participation and referred back to the Prolific survey platform to confirm their participation. Once both platforms had confirmed their participation, their remuneration was authorised.

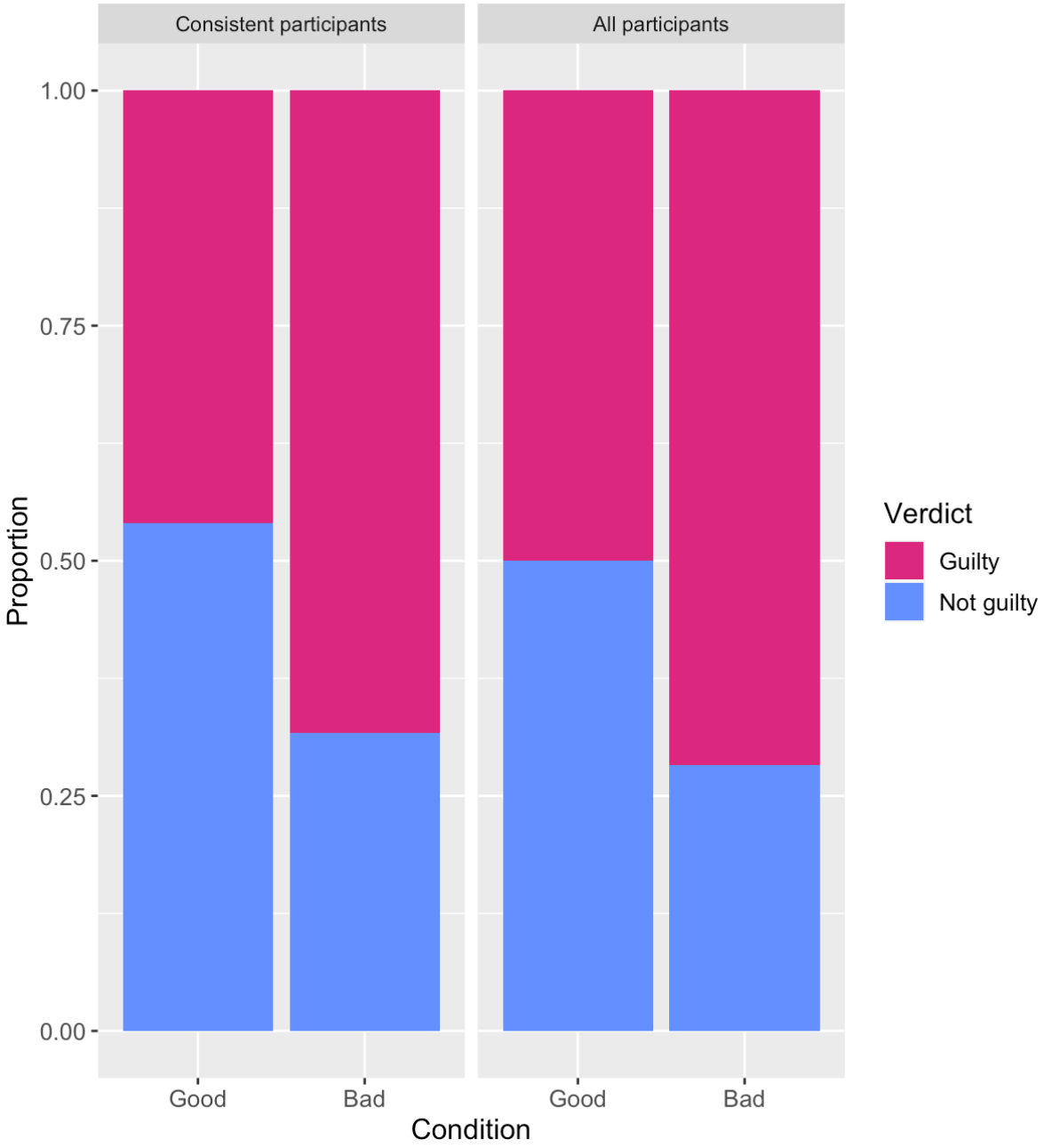
### 4.2.2 Results

The overwhelming proportion of participants (95%) correctly identified the issue in the case when asked on completion of the survey. A small percentage (5%) preferred the intuitively appealing, but incorrect, issue of failing to clear up after a dog. No participants preferred the issue of reasonable excuse. The great majority (91%) also provided a legally sustainable verdict according to the issue. Of the text responses of those who gave a legally unsustainable verdict, three said that they had mistakenly chosen the wrong verdict, and six appeared not to have understood the nature of the task. Considering the reasons given for the verdicts, most participants (93%) had linked their reasons to the issue of whether or not the area was public or not. The reasons of a small minority (7%) had focussed on other issues, such as whether the behaviour was wrong or antisocial.

The findings suggested an influence of character on the determination of issues, and correspondingly on verdicts. Participants in the good character condition were less likely to

find the accused guilty, and were correspondingly more likely to decide that the area in question was private than those in the bad character condition. The proportions of participants who gave legally sustainable verdicts is shown in Figure 1. A similar pattern was evident even where the responses of all participants were considered, also shown in Figure 1.

*Figure 1. Proportions of participants verdicts by condition for both those participants who gave consistent verdicts and for all participants in Study 1.*




We conducted a Fisher's Exact Text (two-tailed) on the cell counts for consistent participants and found that there was a significant association between character and verdict ($p = 0.037$, OR = 2.50, 95% CI [0.99, 6.58]). A very similar result was found when the same analysis was performed on all of the participant data ($p=0.040$, OR = 2.51, 95% CI [1.02, 6.41]).

### 4.2.3 Discussion

#### 4.2.3.1 Summary

Experiment 1 confirmed our hypothesis that, contrary to accepted norms of legal decision making, character would have an influence on the final decision in circumstances where the issues that determined that decision were sufficiently ambiguous that a reasonable decision maker could settle that issue either way. The factual issue of whether an ambiguous area of land was public or not public *prima facie* should not have been influenced by the character of the accused because the issue bore no justifiable legal relationship with character. However, we found that issue, and all of the other information that participants disclosed about their thinking process also appeared to be influenced by the character manipulation.

The normative legal expectation in these circumstances would be that the decision maker would determine the issue of whether the area in question was public or not entirely uninfluenced by the character of the accused, and the outcome of that issue would determine the verdict because there was only one issue in the case. In terms of the decision makers reasons, these should reflect the generalisations relied upon by the decision maker in making the logical leap or inference from the evidence in the case to their factual finding. Superficially, the participants' responses gave the impression that that this was what they had done. Viewed in isolation, almost every participants' individual response appeared to be a legally sustainable decision that complied with those normative expectations.

However, the statistical analysis suggested that in accordance with our hypothesis, a large proportion of participants seemed to have done something more sophisticated. Given that character was manipulated, it appeared that they had started by considering the verdict

that they were more sympathetic to. Those in the good character condition would have been more sympathetic to the defendant and inclined to acquit, while those in the bad character were more inclined to convict. The most plausible explanation for the similar very close association with the resolution of the issue in the case was that a significant proportion of participants were working 'backwards'. That is, once they had determined their preferred verdict, they considered the ambiguity of the issue and the reasons that they might put forward to explain it. Recognising it was possible to credibly resolve the issue and give reasons in accordance with their preferred verdict, many would then give these and finalise their decision.

The alternative explanations that this phenomenon is caused by limitations of cognitive capacity or participants switching to a more heuristic System 1 processing approach do not seem as plausible. For one, it is not clear what heuristic such participants are using. It could not simply a heuristic to favour the verdict depending on character because these participants' behaviour is much more sophisticated than this, given that they also determine the issue and give reasons that are consistent with their preferred verdict. For a second reason, if the 'heuristic' said to be in play is that many participants are choosing the verdict depending on character and then determining the issue and giving reasons to match that verdict, then this behaviour is more cognitively challenging than the normative legal expectation (given that it requires extra considerations in addition to the issue > decision > reasons procedure). Given this additional complexity on top of the normative expectation, it seems inappropriate to label this behaviour as a heuristic. Correspondingly, the behaviour seems unlikely to be either a product of cognitive limitations or a shortcut.

#### 4.2.3.2 Limitations

Our experiment was of course not entirely comparable to the real-life context of magistrates' court decision making. Though many magistrates are lay magistrates, lay magistrates adjudicate in panels of two or three. Thus the pre-decision step of deliberation

was absent and may have made a difference to the outcome. The age range of our subject pool covered all those entitled to apply to be magistrates from 18 to 65, but in practice there have historically been very few lay magistrates aged under 40. Furthermore, lay magistrates will have general experience and a degree of training. That said, there is a limited degree to which adjudicatory skills can be taught, in part because many of the processes are poorly understood, and in part because adjudication tends to consist of applying existing skills to a legal context (Posner, 2008, p. 118). As such, the skills conveyed in judicial training are largely tacit. It would clearly be beneficial to replicate this experiment with experienced adjudicators, but evidence from lay participants provides some support for our hypotheses, not least because research on lay participants tends to be replicated when conducted with professional adjudicators (Hirsch, 2003, p. 601; Kelman et al., 1996, p. 303; Leibovitch, 2016; D. Simon, 1998, pp. 33–34).

#### 4.2.3.3 Further research

Given Study 1 was an intentionally most simple and constrained paradigm with only one issue, we considered it appropriate thereafter to seek to replicate the effects using a more complicated paradigm. The complexity of the paradigm could be increased along one dimension by adding to the number of issues for participants to consider. In addition, complexity could be increased along another dimension by lengthening the chains of reasoning. Thus, whereas the longest chain of reasoning in Study 1 was evidence > issue > verdict > reasons, in the next study one chain could also incorporate the effect of legal precedent. While this would not increase the complexity of the *prima facie* normative legal task that participants were expected to follow, it would increase the complexity of the task for those participants seeking to take account of extralegal information and to obscure this behaviour. Finally, given that Study 1 demonstrated the existence of the phenomenon in a criminal context, our next study would seek to show the phenomenon in a civil context.

## 4.3 STUDY 2

In study 2, we sought to replicate the phenomena observed in study 1, but in both a different and more complicated context. This time we used an example from a civil rather than a criminal context, namely the Employment Tribunals of England and Wales. Furthermore, while the issue that participants were required to determine in study 1 was a factual issue, in study 2 we chose two somewhat different issues: firstly the issue of which previous legal precedent was most analogous to the facts in the study; and secondly the issue of the interpretation of a contract. Again, we manipulated the character of the parties, but this time in a different way because there were now two putative parties given the civil context.

One of the reasons for the choice of the Employment Tribunals as an example of a civil law adjudicatory context was again the role of lay adjudicators. In the Employment Tribunals, decisions will generally be taken by a panel of three individuals, of whom only the chair is legally qualified. The wing members have experience of the workplace context: generally an individual with experience of management and an individual with trade union experience. Given that the participants in the survey would be laypeople, the desire was for an experimental context that primarily involved lay decision makers.

We chose a case with limited legal precedent for the same reasons as before. This was a claim for deduction from wages pursuant to sections 13 to 27 of the Employment Rights Act 1996. In summary, the contract of the employee had ended with the employee owing money to the employer and the employer had sought to deduct the amounts owing from the employee's outstanding expense claims. Deduction from wages cases tend to be relatively straightforward and often settled well in advance of being listed for trial.

For the first issue for participants to determine, we presented two relevant precedents that were both reasonably analogous to the scenario chosen for the survey, but that had opposing implications for the parties: one that ruled that expenses could be considered as wages (*LB Southward v O'Brien* [1996] IRLR 420), the other ruled that expenses were not

wages (*Mears Ltd v Salt* [2012] UKEAT 0522_11_0106). The choice of which precedent was most analogous to the facts in the survey influenced the issue of whether the sums that were deducted could be considered wages.

The second issue for participants was the interpretation of a contractual clause and whether the clause permitted deductions from wages. According to the applicable law, deductions from wages were only permissible if the contract provided for this.

On the basis of the two issues in the case, participants could only find for the employee if they were satisfied to the relevant standard that (1) the sums in question were wages, and (2) the employment contract did not permit deductions. Correspondingly, participants could find for the employer if (1) the sums were not wages, or (2) the employment contract did permit deductions.

Neither of the two issues was linked to the fairness of the deductions. According to the law, the fairness or otherwise of the deductions was not a factor that should influence the Employment Tribunal. Rather, the task for the Employment Tribunal was to decide the formal legal issues set out. To provide an extralegal factor as an experimental manipulation, the background facts were changed between conditions. In the condition designed to provoke sympathy for the employee, the employee was unable to pay back an outstanding loan for training because the employer had fired the employee and replaced him with somebody on a lower wage. In the condition designed to provoke sympathy for the company, the employee did not pay back the loan for training because he had found a better job elsewhere due to developing better skills thanks to the company's loan for training.

In accordance with the theory outlined previously, we hypothesised that the extralegal background information that should have been legally irrelevant would again have an influence on the outcome of the trial, and correspondingly on the determination of the issues, the comparison of the analogies, and the reasons given by the participants. In relation to the issues, we predicted that participants presented with the scenario designed to provoke

sympathy for the employee would disproportionately choose the single combination of issues that would favour the employee whereas participants presented with the scenario that provoked sympathy for the employer would disproportionately favour the other three combinations of issues.

## 4.3.1 Method

### 4.3.1.1 Participants

One hundred and twenty-three participants recruited using the online survey platform Prolific completed the survey (79 females, 42 males, other 2; aged 39 to 69, M age 50.8, SD 7.1; 7% were students; 39% in full time employment, 27% in part time employment, 34% unemployed or other) and were paid £1.50 for their time. The sample size chosen had 80% power to detect an effect size with OR = 3.16 in a two-tailed Fisher's Exact test. To better reflect the demographic of those who sit as Employment Tribunal members, participants were selected on the basis of being aged 40 or above, in addition to having British nationality and being resident in England and Wales. This also reduced the proportion of students in the sample. The study was conducted in accordance with approval obtained from UCL's Research Ethics Committee (CPB/2014/006). Informed consent was obtained from each individual in advance of participation by providing them with study information and a consent form to agree to.

### 4.3.1.2 Design

As with Study 1, we used a between participant design with two levels in which all participants viewed a single set of materials but subject to a manipulation as previously described where the employee was viewed favourably and the employer unfavourably or vice

versa.

#### 4.3.1.3 Materials

As with Study 1, the materials in Study 2 were a realistic transcript of proceedings in the Employment Tribunals of England and Wales, shown to participants using the Qualtrics survey platform. Participants read the transcript which amounted to an introduction to the case and the issues presented by counsel for the claimant and agreed to by counsel for the respondent employer, followed by a record of the evidence presented by the claimant, followed by the evidence of a foreman employed by the respondent. At the conclusion of the evidence, participants read the summing up by counsel for both parties and then they read a summary of the law by the legally qualified Employment Judge.

This time, there was no dispute on the facts. Both parties' evidence was essentially consistent that the employee was provided with a loan by the employer to qualify to drive a 'telehandler' (a large industrial forklift used on building sites) and was subsequently appointed to undertake such work. The employee was also given a flat rate 'mobility allowance' that was in excess of the actual amount that it cost him to get to work and which the employer sought to withhold to compensate for the outstanding loan that had not been paid off at the point that the employee's contract ended. The first issue for participants to determine was whether the mobility allowance amounted to 'wages' as the employer could only lawfully deduct sums from wages, not expenses. Relevant to this first issue was the question of precedent. Participants were invited to consider two precedents: one that implied that payments in excess of actual expenses amounted to wages, the other that payments in excess of actual expenses did not amount to wages. A finding that the sums amounted to wages favoured the employee because he could only bring an action for unlawful deductions from wages if the sums in fact amounted to wages. The second issue for participants to determine was whether the employee's vaguely worded contract permitted deductions to be lawfully made from wages. If the contract did not allow deductions to be made from wages,

then the employers' deductions were unlawful and the employee would win the case. By contrast, if the contract did allow deductions to be made from wages, then the employer would win the case.

To create an extra-legal manipulation that was legally irrelevant to the issues the participants were asked to decide, the background information was changed between conditions. In the condition designed to invoke sympathy for the employee, the employee was unable to pay back the loan for training because the employer had fired him due to finding another employee prepared to drive the telehandler at a lower cost. In the condition designed to provoke sympathy for the employer, the employee had not repaid the loan as he had left his job due to finding a better paid position elsewhere thanks to his new telehandler qualification. When the Employment Judge summed up the legal position, participants were reminded that the issues to be determined were whether the payments were wages or not and whether the contract allowed deductions and that the fairness or otherwise of the deductions was legally irrelevant. Participants were also reminded of the law before they made their final decision. All participants read the same materials save for the character manipulation. At appropriate points in the transcript, portions of the evidence were altered to be consistent with the condition.

#### 4.3.1.4 Measures

Participants were asked to indicate their responses to a number of questions. In relation to the first issue, they were first asked which of the two precedents was more similar to the facts in the instant case: (1) *Southwark v O'Brien* (that an allowance can be expenses even if employees are making a profit because the allowance is in excess of actual expenses incurred); or (2) *Mears v Salt* (an allowance can amount to wages if it is not linked to actual expenses incurred). Participants were then asked to indicate in response to the first issue whether they thought that the employer's mobility allowance was (1) wages, or (2) expenses.

Participants were next asked to indicate for the second issue whether they thought that the relevant clause of the employee's contract: (1) allowed the employer to make deductions for training, or (2) did not allow the employer to make deductions for training.

Then participants were asked, according to the law, who they found in favour of: (1) the employee, or (2) the employer. Finally, they were asked to explain their decision.

#### 4.3.1.5 Procedure

As before, participants were recruited online and participated in the survey in a place of their choosing, using their own device. On referral from the Prolific platform, participants were provided with the study information form. They were next asked to complete the consent form comprising a number of statements to which an affirmative answer was required in order to participate in the survey. Their anonymous user identification was collected to enable subsequent matching of demographic data without compromising the participants' anonymity. They read brief instructions outlining the nature of the task and what they were to expect.

Participants were randomly assigned to one of the two conditions by the Qualtrics survey platform and thereafter viewed a single version of the two versions of the case described above, each of which was identical other than some of the background information that supported the condition designed to invoke sympathy for the claimant employee or the respondent employer. Participants read the outline of the case set out in the summary by the claimant's counsel, the transcripts of examination-in-chief and cross-examination of both the claimant and the respondent's foreman, the closing summaries by counsel for both parties, finally followed by the legal advice from the Employment Judge.

Following completion of the review of the evidence, submissions, and advice, participants were then given final instructions reminding them of the task that they were

asked to complete where they were again reminded that how the employment contract ended was not legally relevant. They were then asked to complete the measures discussed in the previous section.

After participants had indicated their responses to the measures, they were asked to select from a number of correct and incorrect statements about the issues in the case. Next, those who had provided a legally inconsistent response (such as determining an issue inconsistently with their final decision) were asked to explain why they had done this.

After reviewing the transcript, participants were asked to complete the measures described in the previous section and the additional measures, if applicable.

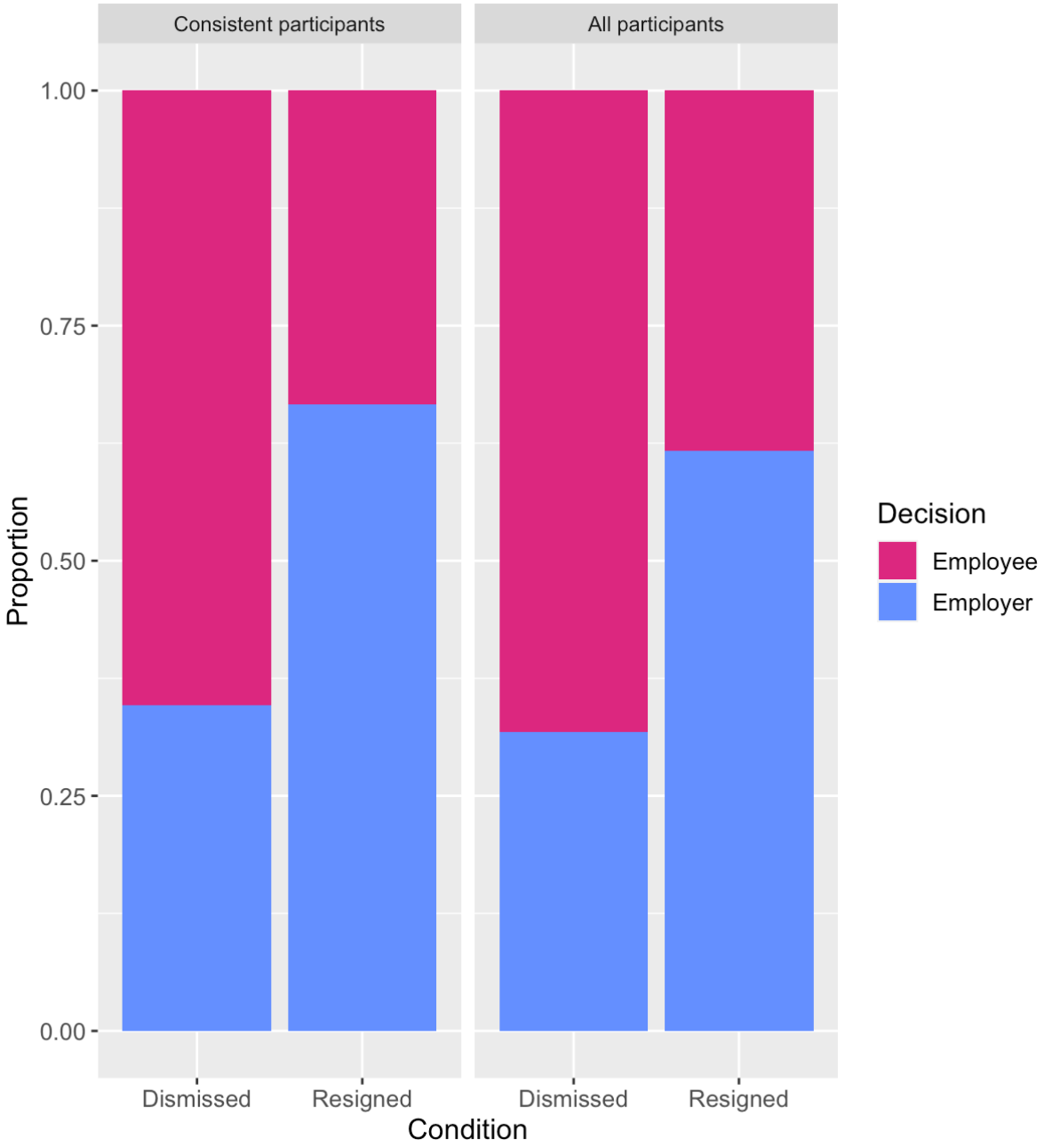
After completing the survey, participants were thanked for their participation and referred back to the Prolific survey platform to confirm their participation. Once both platforms had confirmed their participation, their payment of £1.50 was authorised.

### 4.3.2 Results

Of the 123 participants, 100 (81%) provided a legally sustainable verdict. After completing the survey, 31% correctly identified the two issues in the case. However of particular note was that a greater proportion, 58%, additionally indicated that the reasonableness of making the deductions was an issue in the case, despite the instructions on the law and despite the absence of a measure that would correspond with such an issue at the point when participants made their final decision. Other incorrect combinations of issues were indicated by some participants, but with a very low frequency, namely a maximum of 2%. Notwithstanding responses to the post survey question on the issues in the case, relatively few participants actually mentioned extralegal factors such as fairness in the reasons justification their decision (15 out of 123 or 12%).

In terms of final decisions, a greater proportion of participants found in favour of the employee in the condition where he was dismissed which was designed to invoke sympathy for him and conversely a smaller proportion found in his favour in the condition where he resigned that was designed to invoke sympathy for the employer, despite sympathy being legally irrelevant to the issues. This was the same pattern for both all participants as well as those participants who gave legally sustainable responses: see Figure 2.

*Figure 2. Proportion of participants' decisions in favour of employee and employer according to whether the employee was dismissed or resigned from Study 2.*

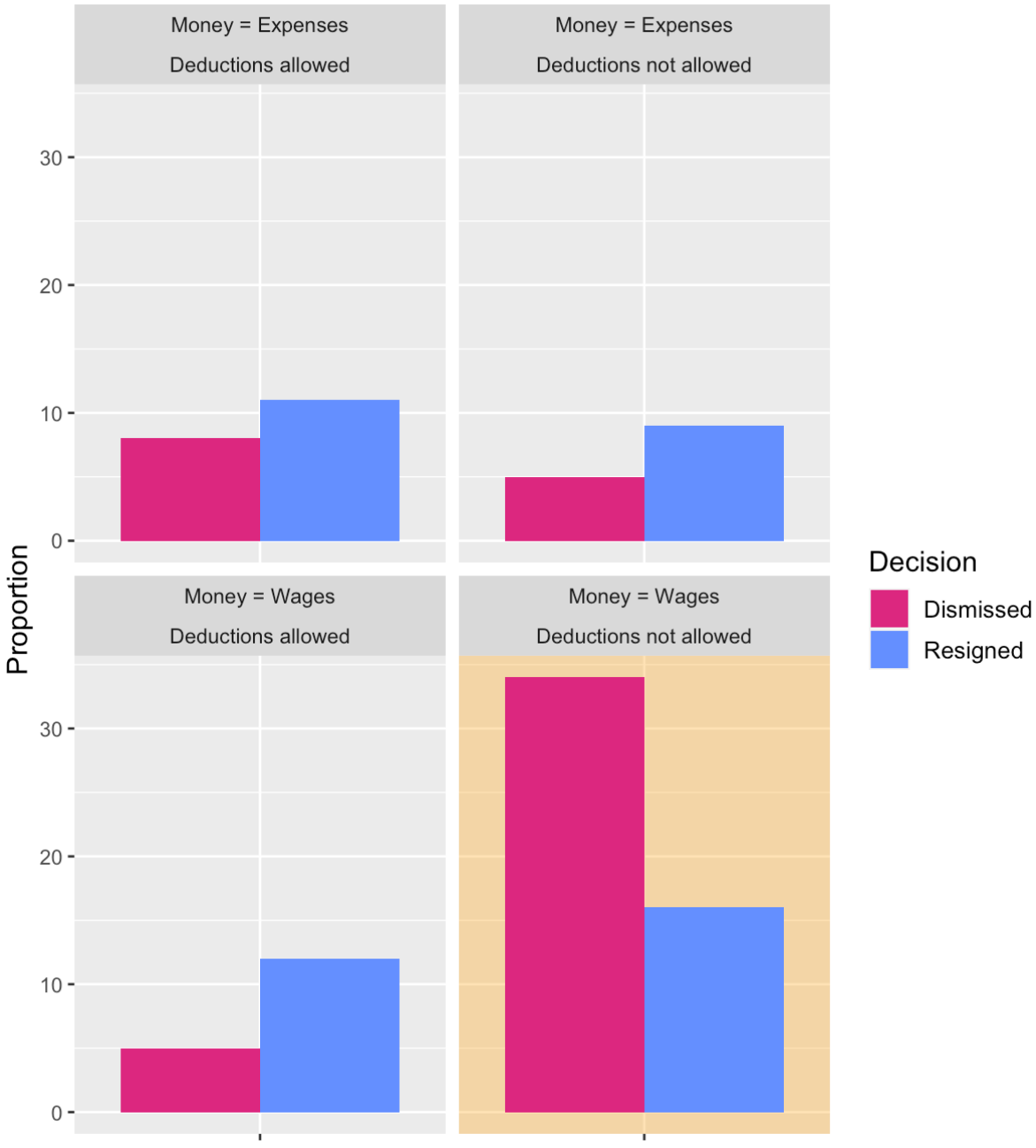



A Fisher's exact text was performed on this data which confirmed that the findings were significant, both for all the data: p = 0.00078, OR = 3.42, 95% CI [1.73, inf]; and for

those participants who gave legally sustainable responses: p = 0.0013, OR = 3.72, 95% CI [1.73, inf]. This indicated that notwithstanding the fact that the reasonableness of the deductions was not an issue in the case (and participants were given no opportunity to indicate their view as to the reasonableness of the deductions), this had nonetheless influenced final decisions.

Turning to the issues, the data suggested that there was also a relationship between the conditions and participants' findings on the issues such that participants interpretations of the precedent and the contract favoured the employee in the conditions designed to invoke sympathy for him, and favoured the employer in the conditions designed to invoke sympathy for the employer. This pattern was visible from both all participant data and for those participants who gave consistent verdicts. Given that there were two binary issues, there were four possible combinations of findings by participants. Three of these combinations favoured the employer and only one favoured the employee. In the condition where participants were advised that the employee had resigned, participants were much more likely to choose one of the 3 possible combinations of issues that favoured the employer. By contrast, where participants were advised that the employee had been dismissed, they were much more likely to settle on the single combination of issues that favoured him. This is shown in Figure 3, for those participants who gave legally sustainable decisions. The first three quadrants favour the employer, and the final quadrant (highlighted) favours the employee.

*Figure 3. Issues preferred by participants depending on condition in Study 2.*




Log linear models were built to assess whether there was any statistically significant relationship between the manipulation and participants' responses. One model (H0) was built containing parameters for the issues and for the manipulation to represent the hypothesis that

there was no relationship between the manipulation and participants' responses. A second model (H1) was built with additional parameters for the outcomes that favoured one side or the other and the interaction between that parameter and the manipulation. Table 1 summarises the models:

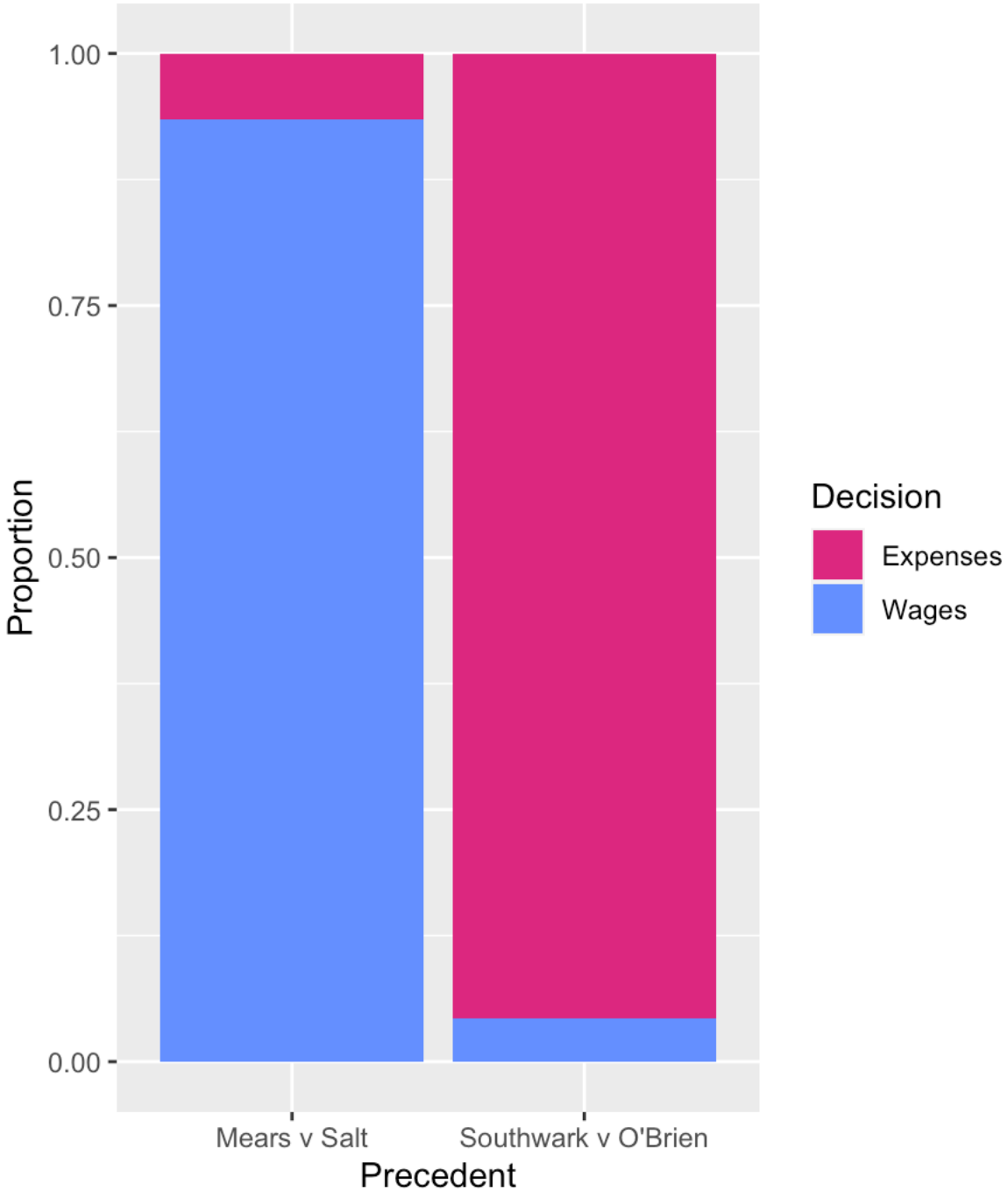
*Table 1. Summary of models used to assess Study 2*

| **Model** | **$H_0$** | **$H_1$** |
|---|---|---|
| **Null deviance** | 30.18 on 7 df | 30.17 on 7 df |
| **Residual deviance** | 13.68 on 4 df | 1.37 on 2 df |
| **p on residual deviance** | 0.0084 | 0.50 |

The residual deviance of the models indicated that H0 was a poor fit to the data compared to H1. The superior fit of H1, the model with additional parameters, was confirmed statistically with a $\chi^2$ comparison of the models: $\chi^2(2) = 12.307$, p=0.0021. For H1, the odds ratio for the interaction was 2.70 [CI 1.30, 5.72], indicating that participants were significantly more likely to prefer an outcome on the issues that favoured the party that matched the experimental manipulation, even though this the experimental manipulation was legally irrelevant to the issues the participants were required to determine.

Also of relevance was the relationship between participant's views on which previous legal precedent was more analogous and their findings on the issue of whether the payments were wages or not. In addition to the above finding that participants' views on the issue of whether the sums paid were wages or not, participants' views as to which analogy was most similar to the facts of the case were also extremely closely tied to their decision on this issue, see Figure 4.

*Figure 4. Participant's assessment of appropriate analogy depending on decision on the issue of whether money was wages or expenses in Study 2.*




Unsurprisingly, a Fisher's exact test performed on this data showed a highly

significant relationship: $p < 0.001$, OR = 228, 95% CI [52; 3,108]. In the light of the close relationship between the issue of whether the sums were wages or expenses and the manipulation, this indicated that there was also a very close relationship between the experimental manipulation and participant's views on the analogy.

### 4.3.3 Discussion

#### 4.3.3.1 Summary

In Study 2 we replicated the patterns seen in Study 1, but in a slightly more complicated scenario. Both studies entailed lay adjudication, but whereas Study 1 was in a criminal context, Study 2 was a civil context. In addition to doubling the number of issues for participants to consider, we also increased the complexity of one of the issues by lengthening the chain of reasoning that participants would be required to follow. In Study 1 the putative chain of reasoning was facts > issue > decision > reasons, in Study 2, the putative chain of reasoning for the issue of whether the sums paid were wages or expenses was facts > analogy > issues > decision > reasons.

According to legal normative expectation, neither the decision, nor the assessment of the issues, nor the drawing of the analogy should have been affected by the experimental manipulation, and similarly the reasons should not have referred to any aspects of the manipulation. However, the manipulation put moral pressure on the decision maker because it seemed very unfair to the employee to require him pay back a loan in the scenario where he had been dismissed by the employer and similarly it seemed very unfair for the employee to not pay back the loan when he had used the benefit of the loan to secure another position elsewhere.

The results suggested that in accordance with our predictions, and contrary to legal

normative expectation, the manipulation had had an influence on all aspects of the decision making and reason giving. Participants in the condition designed to invoke sympathy for the employee ultimately found in his favour. What was striking was that in addition to the final decisions being influenced by the manipulation, the associated chains of reasoning invariably gave the appearance of being coherent with those decisions. Thus, a participant in the condition sympathetic to the employee was statistically much more likely to decide that (1) the sums were wages, and (2) the employment contract did not permit deductions (the only combination of findings on the issues that would permit a finding in favour of the employee). In turn, such a participant would also prefer the analogy of *Southward v O'Brien* (which implied such sums were wages). And, in accordance with previous research (Braman, 2006, p. 310; Liu, 2018, p. 96), participants rarely mentioned extralegal factors such as reasonableness or fairness in their reasons, generally preferring to explain their thinking by reference to the legal issues.

Thus, if one examined a single participants' responses in isolation, they would appear to have followed the analytical approach expected by normative legal expectation, eg in the order facts > analogy > issue > decision > reasons. However, the statistical analysis with the benefit of counterfactuals and a sufficiently large sample size revealed that many participants appeared to have worked backwards from the outcome that they preferred, ie they impermissibly took into account fairness or reasonableness when deciding which decision they preferred, and then subsequently worked out a plausible inference process to justify that decision.

As before, such behaviour does not seem to fit with theories that consider such phenomena as a shortcoming, mental shortcoming, or heuristic because the process that participants seemed to follow appears to be more cognitively demanding that if they had followed the analytical approach envisaged by standard legal theory. Ie, participants seem to be taking account of an additional factor, namely fairness or reasonableness, and participants also seem to be assessing which inference procedure would most plausibly account for their preferred outcome.

#### 4.3.3.2 Limitations

One query about Study 2 arises out of the relatively large proportion of participants who apparently mis-identified one of the issues that they were supposed to be deciding as including the reasonableness of making the deductions. This could be interpreted a number of different ways. One interpretation is that, due to their relative lack of legal experience, they misunderstood the instructions and legal guidance and were incorrectly applying the legal test, thereby leading to the empirical findings that we observed. If this was the case, this might reduce the evidential support for our hypothesis somewhat, even if it would not necessarily provide evidence in favour of the alternative irrationality or cognitive capacity hypotheses. However, there is reason to question whether this is the most appropriate interpretation. For one, these participants correctly identified the other issues and if they did in fact believe that they were assessing the reasonableness of the deductions, we would have expected this to be reflected in the reasons for their decisions. However, we did not: relatively few participants mentioned reasonableness when explaining their thinking and many of these mentioned reasonableness quite obliquely. Another interpretation of these responses is that participants wanted to take into account reasonableness, but given that their decision was constrained by the law, they did so covertly rather than overtly. This would be in accordance with our hypothesis. However, the most plausible explanation is probably that the question was simply ambiguously worded, meaning that participants interpreted it as asking whether their decision should be reasonable overall, rather than whether they should take into account extralegal factors. Again, such an interpretation would be compatible with our hypothesis.

#### 4.3.3.3 Further Research

The findings from Studies 1 and 2, together with other previous empirical research, suggest other means of distinguishing between the theories that seek to explain the effect of extralegal information on legal decision making. In particular, we know that adjudicators assess the impact of evidence on their mental model on a continuous basis and as each new piece of evidence is presented rather than waiting for a particular point to do so (such as when the evidence is considered complete) (Holyoak & Simon, 1999; N. Pennington & Hastie, 1988; N. Pennington & Reid, 1993; D. Simon, 2004; D. Simon et al., 2001). By the same token, extralegal information seems to bias outcomes and information associated with those outcomes as soon as it is made available (Holyoak & Simon, 1999; D. Simon, 2004; D. Simon et al., 2001). Thus far, character manipulations have tended to be single and between participant, leaving open an opportunity to explore the effect double, within participant, manipulations and how these illuminate the plausibility of different theories. Given that irrationality type theories assume that extralegal information is taken into account due to cognitive limitations or reliance on heuristic shortcuts, such theories would seem to predict that a second character manipulation ought not to have any further effect on participants' behaviour, given that it would simply be increasing the complexity of the decision further. By contrast, the theory that we have been exploring that extralegal information is taken into account in order to further the adjudicator's preferences would imply that a second, opposing, character manipulation should have the reverse effect on the adjudicator's decision (and the chains of reasoning linked to it).

## 4.4 STUDY 3

In Study 3 we sought to examine the effect of changing character during the presentation of evidence in a trial. Studies 1 and 2 demonstrated that character caused participants apparent assessments of issues unrelated to character to change, which we theorised was because the determination of these issues in a particular way was necessary for

participants to make a final determination in favour of the party with which they sympathised. Previous adjudicatory research into adjudication has indicated that participants' assessments of the issues change dynamically during the presentation of evidence as material that that affects those issues is presented (Holyoak & Simon, 1999; N. Pennington & Hastie, 1988; N. Pennington & Reid, 1993; D. Simon, 2004; D. Simon et al., 2001). For example, Simon et al found that participants' assessments of the reliability of a witness identification changed after incriminating or exonerating DNA evidence was presented that supported or undermined the witness identification (D. Simon, Krawczyk, et al., 2004). This pattern would tend to be expected by standard accounts of rational decision making (Lagnado & Gerstenberg, 2017). Simon et al found some evidence that extralegal information was also taken into account in a dynamic way, influencing participants' putative assessments of issues as soon as it was presented, rather than when the presentation of all the evidence was complete (Holyoak & Simon, 1999; D. Simon, 2004; D. Simon et al., 2001). The effects seemed relatively robust and were manifested not only when participants were told that they were determining a legal case, but also when the same information was presented as a memory test or a comprehension test (D. Simon et al., 2001).

Previous research has primarily manipulated only the effect of character between participants such that each participant saw only one version of the materials. We were therefore interested in examining the effect of a dual character manipulation such that participants initially assessed the evidence in the light of a character manipulation, but subsequently reassessed the evidence in the light of a manipulation in the opposite direction. In the light of the developing theory, we predicted that the first character manipulation would influence the participant's determination of issues with which it had no logical or legal relationship in the predicted direction and the second, opposing, manipulation should influence the participant's assessment in the opposite direction. To do this, we developed a paradigm based on the approach previously used by Simon and his collaborators whereby at a point during the presentation of evidence, participants were invited to make a 'preliminary assessment' of the evidence that was confidential to the participant while being given the impression that further information was due to be presented. Our paradigm was loosely based

on the Jason Wells case, a relatively complicated scenario that included both issues that might be influenced by character, as well as issues that arguably ought not to be so influenced. We similarly included both issues that bore a logical relationship with character as well as issues that did not. However, whereas Simon's case of Wells was set in a criminal context, we based our scenario on a workplace adjudication so as to replicate the type of decision that might be undertaken by a lay decision maker, as well as avoiding issues to do with presenting character evidence in a way that would be unlikely to be admissible in a criminal context.

Participants were asked to determine a workplace disciplinary case of alleged gross misconduct. The facts were that Tom Clarke, an employee of Paragon, had attended an industry awards event after normal office hours. At the awards, he was alleged to have drunk too much and been disorderly, behaviour which was alleged to amount to gross misconduct. The first character manipulation was a confidential reference said to have been provided by Tom's line manager and shown to participants before they viewed any of the materials. This was either a glowing reference or a fairly damning one. The second character reference was a statement from a member of Paragon's administration admitting that the first reference that was provided was not genuine and had only been provided because she had been blackmailed into providing it. She also attached what was said to have been the genuine reference. Where a good character reference had initially been provided, the blackmail was said to have been undertaken by Tom, and where a bad character reference had initially been provided, the blackmail was said to have been undertaken by Tom's line manager. There was also a third condition with no character manipulation at either the beginning or end. Thus participants saw either a good, then bad, reference or vice-versa, and the explanation for the new reference reinforced the intended effect of the second reference, or they saw no character references at all. There were then 6 issues for participants to determine, 3 of which could plausibly be linked to the character manipulation, and 3 which could not. For example, An issue which could be linked to character was whether or not Tom was badly behaved at the awards. An issue which was more difficult to link to character was the interpretation of the employment contract, and whether the ambiguous wording could be interpreted to encompass being drunk within gross misconduct.

We predicted that for participants in the two groups who were subject to a character manipulation, their views on the 3 issues linked to character should be influenced by both character manipulations, and that participants' views on the 3 issues not linked to character should also be influenced by both character manipulations. The R statistical analysis was preregistered with Open Science Framework.

## 4.4.1 Method

### 4.4.1.1 Participants

One hundred and fifty participants recruited using the online survey platform Prolific completed the survey (100 females, 48 males, 2 other; aged 19 to 64, M age 31.8, SD 10.8; 14% were students; 61% in full time employment, 17% in part time employment, 22% unemployed or other; all were UK nationals of whom 98% lived in England and 2% lived in Wales and 96% were born in the UK). A range of simulations were undertaken to understand the effect of the sample size on the power of the study. For instance, for the comparisons of primary interest (the relationship between character and the measures either expected or not expected to be influenced by character), assuming a standardised residual variance of 0.7 and a standardised variance of 0.3 for a random intercept for participant, the sample size was calculated to have an 80% power to detect an effect size of r2=0.3. Due to an undisclosed information policy by Prolific, some participant data had expired by the time the analysis was undertaken and for that reason could not be used. Participants were paid £1.00 for their time. The study was conducted in accordance with approval obtained from UCL's Research Ethics Committee (EP/2018/005). Informed consent was obtained from each individual in advance of participation by providing them with study information and a consent form to agree to.

#### 4.4.1.2 Design

For Study 3 we used a between participant design with one independent variable and three levels in which all participants viewed a single set of materials but where two of the groups were subjected to two character manipulations (either good then bad, or bad then good) and the control group was subjected to no character manipulation.

#### 4.4.1.3 Materials

Participants were told that they were being asked to play the role of an independent arbitrator who has been asked to decide a workplace disciplinary case and that they should make their decision on the basis of the evidence and arguments from both sides. Participants were provided with the text of a disciplinary charge that alleged the employee, Tom Clarke, was intoxicated at work during an industry awards presentation, a charge that amounted to gross misconduct.

Before the evidence, participants in the two conditions subjected to a character manipulation were shown a confidential character reference regarding Tom which either suggested he was of good character or bad character. Those in the control condition were shown no character information.

Participants were provided with the evidence in the case which included a relevant extract from the company's disciplinary policy. This included an ambiguous clause that defined gross misconduct, but which did not explicitly refer to being intoxicated.

Participants were told that Tom was a full-time employee, having completed his probation. The evidence against him was from a client of the company who attended the awards presentation and said that some attendees took advantage of the free bar and that Tom was loud, boorish, and disruptive, slurring his words, and was so drunk at one point he could

hardly stand. The customer said that Tom was not his main point of contact, but somebody who he had met once or twice at technical meetings. In response, Tom said that he did attend the awards, but said that he was not drinking. He claimed to have only had soft drinks. He said that he had to leave before the main presentation because he needed to go home to look after his children while his wife went to the gym. He admitted to vaguely knowing the customer, but said he thought he must have confused him for somebody else.

Participants were told of a previous precedent where a female member of staff became unwell during an impromptu celebration of a sales target at which champagne was served. The member of staff was sent home in a taxi with no disciplinary proceedings being instituted.

At the conclusion of this evidence, the arguments of the parties were shown to participants, with the respective sides addressing the ambiguities in the issues.

The further evidence was then shown of another character manipulation in the opposite direction of the first manipulation. Thus those participants who first saw a good character manipulation were then shown a second manipulation that undermined the first manipulation and suggested that Tom was in fact of bad character, and vice versa. Those participants in the control condition were again shown no character manipulation.

#### 4.4.1.4 Measures

Participants were asked the same substantive battery of questions at two points in the survey. The first time the questions were administered was after the presentation of the first reference (save in the control condition where the reference was not provided) and the other materials. This was at a point when participants were explicitly told that they were waiting for further material to be presented. The battery of questions was described as a 'preliminary leaning' which was confidential to the participant and that would not be seen by the parties.

As part of these questions, participants were first asked towards which party their preliminary leaning was and were required to indicate a response to a binary question of one party or the other. They were also asked on a scale of 0 to 5 with accompanying verbal descriptions of 'not at all confident', 'reasonably confident', and 'very confident' at the ends and midpoint of that scale.

Then, for each of the seven issues (six plus an attention check of whether Tom was on probation or not), participants were invited to indicate their responses on a scale from -5 to +5, the extent to which they disagreed or agreed. Accompanying the scale at either extreme and then equally spaced were verbal descriptions of 'strongly disagree', 'disagree', 'somewhat disagree', 'neutral', somewhat agree', 'agree', and 'strongly agree'. The issues were (1) whether Tom was on probation, (2) whether Tom left the awards before the presentation ceremony, (3) whether a witness correctly identified Tom, (4) whether Tom was drunk and disruptive, (5) whether attendance at industry awards counted as 'at work' under Paragon's disciplinary policy, (6) whether being intoxicated at work amounted to gross misconduct under Paragon's disciplinary policy, and (7) whether a previous incident where a staff member was not disciplined was similar to the disciplinary charge levelled at Tom. These seven issues were presented in a random order to each participant. The probation question was used as an attention check as Tom was described as a 'full-time employee' who had 'completed his probation'.

The same battery of questions were then presented again immediately after the presentation of the second reference (where applicable) where they were described as a final, public, decision that would be seen by the parties. Rather than being described as a 'preliminary leaning' this was now described as a 'final decision'.

#### 4.4.1.5 Procedure

Participants were recruited online and participated in the survey in a place of their

choosing, using their own device. On referral from the Prolific platform, participants were provided with the study information form. They were next asked to complete the consent form comprising a number of statements to which an affirmative answer was required in order to participate in the survey. Their anonymous user identification was collected to enable subsequent matching of demographic data without compromising the participants' anonymity.

Participants read brief instructions outlining the nature of the task and what they were to expect. They were described as playing the role of an independent arbitrator asked to decide a workplace disciplinary case with the responsibility of deciding the case in favour of the company or the employee. Participants were told that they were not required to have any legal knowledge and that they should use their common sense when making their decision.

Participants were told that there might be some further evidence that would come available. They were advised that they had decided to consider the available evidence and arguments, but not reach a final decision until they had seen the further evidence. On viewing the further evidence, they would make a decision. Participants were told that they would see the evidence, followed by the arguments of the company and the employee. Once they had considered the information, they would be asked to make a preliminary evaluation that would be confidential and not seen by the parties. They were told that after viewing any further evidence, they would then make a final decision that would be shared with the parties.

Those taking part were randomly assigned to one of the three conditions: (1) good followed by bad character, (2) bad followed by good character, and (3) no character manipulation at either stage.

At the outset, those participants in the first two conditions were shown either a good or bad reference whereas those participants in the control condition were shown no reference. Thereafter, all participants were shown the same materials, consisting of a summary of evidence in which the employer's evidence as that a third party identified Tom at the event and described him as drunk. By contrast, Tom denied drinking alcohol at the event and

claimed to have left before the awards presentation itself where the witness described him as being drunk and disruptive. Participants were shown extracts from the employment contract that were ambiguous as to: (1) whether attendance at external events counted as 'at work' and (2) whether being drunk amounted to gross misconduct. Finally they considered a previous precedent where by an employee who had apparently been drunk at work on a previous occasion was not disciplined. Participants read arguments by both parties that addressed and reiterated the issues in the case.

After reviewing the materials and being reminded that further materials might be expected, participants were then invited to give their confidential preliminary leaning where they were asked to indicate their responses to the battery of measures outlined previously, each issue being presented in a randomised order. They were able to refer again to the materials and arguments if they wished to.

Thereafter, participants in the two conditions with a character manipulation were presented with an opposing character manipulation to that shown at the outset. Participants in the control condition were advised that no further evidence would become available.

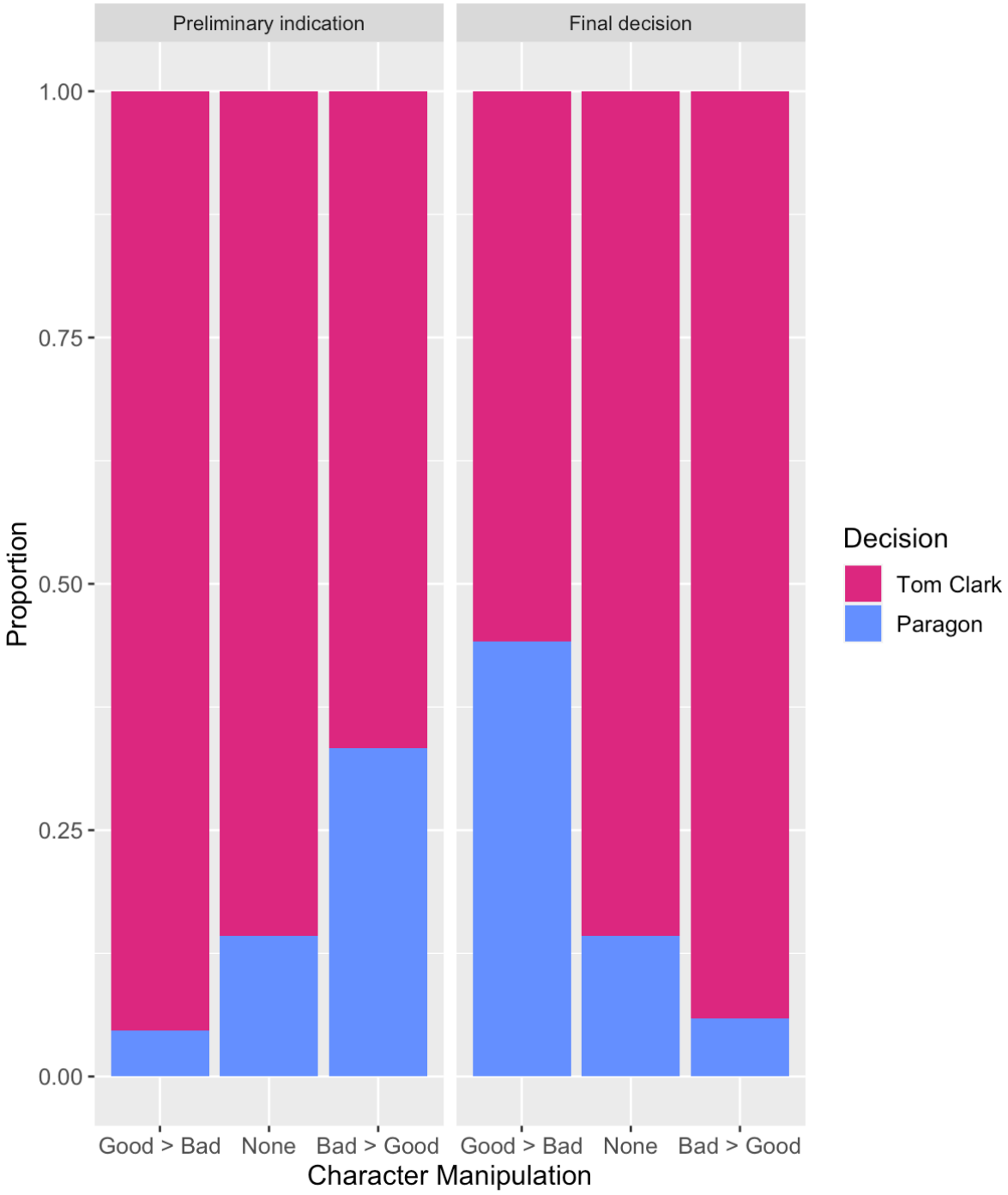
At the conclusion, participants were asked to give their final, public, decision that would be seen by the parties. Participants then completed the same battery of questions which were now described as a final decision. As before, the issues were randomised and participants were able to refer to the materials and arguments.

### 4.4.2 Results

Of the 150 participants, 28 (19%) failed the attention check question by incorrectly indicating that Tom was an employee on probation. This data was excluded from the analysis. Though the final decision was not one of the key metrics relevant to our hypothesis, the participants responses shown in Figure 5 demonstrate that the character manipulation

appeared to be effective. For those in the control condition with no character manipulation, the overwhelming finding was in favour of Tom (24:4) and predictably this did not change between the preliminary leaning and the final decision. By contrast, where participants saw Tom as of good character initially, their preliminary leaning was in his favour (41:2) switching to a more split view (24:19) when it transpired he was of bad character. Similarly, those who thought Tom of bad character initially were similarly split (34:17), but then overwhelmingly in his favour when they found out he was of good character (48:3). A generalised linear mixed effects model (binomial) was built to predict decisions with fixed effects for character and response (preliminary leaning and final decision) and with a random grouping term for participant. Unsurprisingly, this pattern was statistically significant (B=15.3209, z=3.99 p=0.0028, OR=4,506,107 (CI=196.9; 103,121,588,226)). In terms of confidence, the mean confidence response was 3.42 (SD 0.952) on the scale from 0-5, suggesting that participants were quite confident in their decisions. There was no difference in confidence between any of the groups, other than that participants giving a preliminary indication (M 3.21, SD 0.80) were slightly less confident than those giving a final decision (M 3.62, SD 1.05). A mixed effects model containing parameters for group and position confirmed that this difference was statistically significant (B=0.588, df=214.6, t=3.099, p=0.002). This difference was perhaps unsurprising given that participants were specifically advised that they were awaiting further evidence as at the point they gave their preliminary indication.

*Figure 5. Participants' preliminary indications and final decisions according to character manipulation in Study 3.*

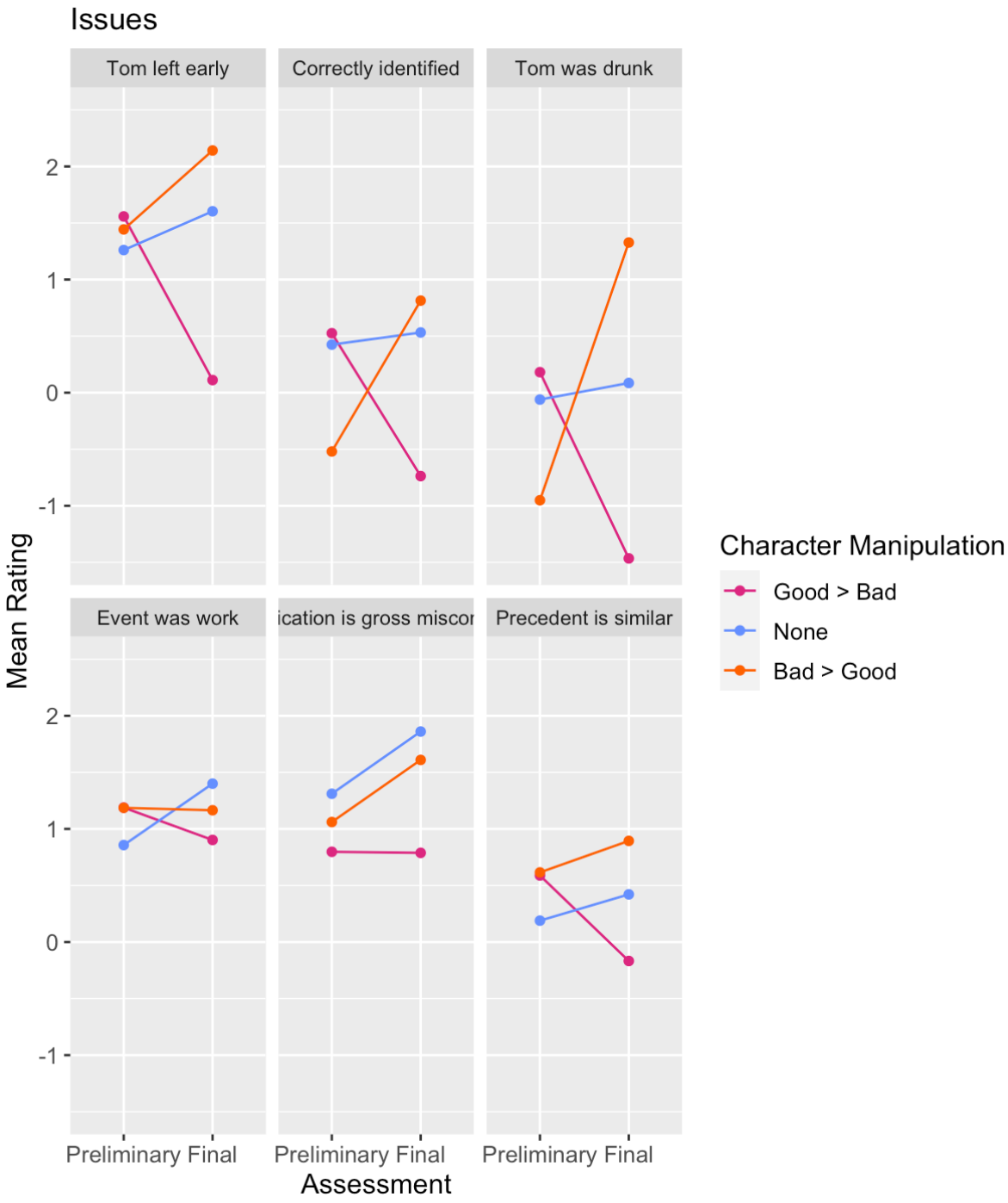



Similar patterns were seen in the overall averages of the measures, but this does not

distinguish between the measures that were expected to move and those that were not.

A more in-depth picture of participants' responses was available by examining their responses to the individual issues as shown in Figure 6. The responses to the three issues predicted to be influenced by character shown in the top row showed the predicted pattern, with the assessment of the issues being influenced by character as would be logically and legally appropriate. Predictably, responses in the control condition where there was no character manipulation showed little change. By contrast, responses to the issues that were not linked to character shown in the second row changed significantly less.

*Figure 6. Participants' responses to the individual issues by character in Study 3.*




To test these responses statistically, we built a linear mixed effects model with fixed

effects for the issues expected to be influenced by character (whether Tom left early, whether he was correctly identified, and whether he was drunk), fixed effects for character and response (preliminary leaning and final decision), a random grouping term for participant, and a random slope for the issue. As expected, this showed a significant difference by character and interaction between character and response. Here, the beta estimate for character was (good v bad) was 1.44 (df=119.0, t=8.48, p<0.001, r2=0.08), demonstrating that character had a big impact on participants' assessment of the issues, favouring Tom when he was of good character and vice versa. Also of relevance was that the beta estimate for the interaction between character (good v bad) and response was -1.36 (df=119.0, t=-2.03, p=0.045, r2=0.02). The mean value of the assessments for good character was lower for the preliminary assessment than the final decision (0.76 v 1.43) and the mean value for bad character was higher for the preliminary assessment than the final decision (-0.01 v -0.70). This suggested that participants' assessments became even more extreme for the final decision, possibly because the second character manipulation was stronger than the first.

In relation to the data that we were most interested in for this experiment, responses to issues that standard expectation predicts ought not to have moved, but that our theory predicted would move, the results were not as expected. We built a similar linear mixed effects model as above for the issues expected not to be influenced by character (whether attendance at an after hours events counts as at work, whether being drunk amounted to gross misconduct under the employment contract, and whether the previous case was similar to the allegations in Tom's case). Again, we included fixed effects for character and response, a random grouping term for participant, and a random slope for the issue. However, contrary to our predictions, participants' responses did not seem to be influenced by character in a statistically significant way. The beta estimate for character (good v bad) was only 0.31, which was not statistically significant (df=590, t=1.50, p=0.13, r2=0.003). Again participants' assessments were less extreme for the preliminary assessment than the final decision (good: 0.86 v 1.22; bad: 0.95 v 0.51), but this did not reach statistical significance (B=-0.81, df=119, t=-1.038, p=0.30, r2=0.004). Nonetheless, what was noteworthy was that there was practically no extralegal effect of character at the preliminary assessment stage, whereas there

was an indication of an extralegal effect at the final stage. While the latter did not reach statistical significance, it does appear consistent with an exercise of caution by participants at the preliminary stage compared to the final stage where there is some evidence consistent with an extralegal effect.

### 4.4.3 Discussion

The empirical findings of Study 3 were more in accordance with normal common-sense expectations of how adjudication should proceed rather than illustrating the types of empirical findings that we have been exploring. The study provided participants with a more complicated task than in previous experiments. The task required determination of both issues that common sense would assume to be influenced by character manipulations as well as issues that common sense would assume not to be influenced by character manipulations. Give our theory and previous empirical findings, we expected that the latter category of issues would in fact be influenced by character. Yet this is quite not what we saw. Rather, participants apparently distinguished between the two categories of issues: those that were linked to character were influenced by character (at both the preliminary indication and final decision stages) and those that were not linked to character were not influenced by character at the preliminary stage, though there was some evidence consistent with a slight effect at the final decision stage.

These findings are not well explained by irrationality or cognitive complexity theories that assume that the non-normative empirical findings are caused by the complexity of the adjudicatory task or lack of cognitive capacity on the part of the adjudicator. Given that the task was more complex than in Experiments 1 and 2, these theories would appear to predict that the extra-legal empirical findings would be more pronounced. But because the empirical patterns did not appear, this experiment appears to provide little or no support for the irrationality family of theories.

Similarly, the findings seem to provide limited support for our developing theory that the non-normative empirical patterns are evidence of a cognitive strategy to secure the outcomes that the adjudicator favours through managing the information available to observers about that inference process. We predicted that the category of issues that ought not to have been influenced by character would have in fact been influenced at the preliminary indication stage, and also influenced at the final decision stage but in the opposite direction following the second, opposite, character manipulation. However, the findings suggested no influence at the preliminary indication stage and only a slight effect at the final stage.

Faced with a situation where neither theory explains the findings well, it seems appropriate to examine the auxiliary assumptions of our theory to see if adjustments to these assumptions can accommodate the findings. Key differences between Experiment 3 and previous research were the dual manipulation, the nature of the measures, the combination of issues that ought to have been influenced by character as well as those that ought not to have been so influenced, and the increased complexity. Of these, the dual character manipulation appears to be of limited relevance because the envisaged empirical patterns were not seen at the stage participants made their preliminary indication, a point where there had only been a single character manipulation. By contrast, the nature of the measures appears potentially relevant. In our previous experiments, the measures were binary so that participants were required to indicate clearly one way or another. In Experiment 3 we adopted a Likert scale as Simon et al had previously. A scale is not obviously a determination one way or another, save at the extremes of the scale. One impact of this is that it would be harder for an observer to say for certain whether a participant's responses to the issues was inconsistent with their final decision. Consequently there would be less pressure for a participant to give a particular response to an issue in order to appear consistent. Along similar lines, the greater number of issues (both those expected to change by character as well as those not expected to change) would also have made it difficult for an observer to discern any inconsistencies between a participant's final decision and their views on the individual issues, possibly also reducing the pressure on participants to manage the information about their decision processes to appear consistent. If these differences do explain the unexpected empirical findings, this might

appear consistent with previous research whereby the empirical patterns we are interested in seem common in experiments where the decision environment is quite tightly constrained (for example our Studies 1 and 2), but less common where the decision environment is more uncertain. For example, Simon's findings in the Quest paradigm were only marginally significant (Holyoak & Simon, 1999, p. 12). An alternative explanation for the unexpected findings of Experiment 3 might be linked to the additional cognitive challenge for participants taking extra-legal information into account. In other words, all participants have to consider the nature of the underlying decision, but those participants taking extra-legal information into account also have to consider whether their decision would appear reasonable to third parties and behave accordingly. This additional task might be straightforward in simple cases, but could prove increasingly onerous as the nature of the underlying decision becomes more complicated. It could be a possibility that there reaches a point where the secondary task becomes too complicated and participants revert simply to considering the underlying decision. Finally, while there were obviously a number of other differences between previous research and Experiment 3, such as its context as a workplace adjudication, these did not seem to be as likely to be material.

## 4.5 STUDY 4

In the light of the findings from Study 3, Study 4 was a pared down version of the paradigm in order to see if greater constraints would cause the same extralegal influences previously identified in Studies 1 and 2. In a more constrained scenario, we predicted that there would be the usual effect of character on the final decision, but accompanied by a greater pressure on participants to show consistency between their final decision and their decision with the issues. Participants were only asked to consider those issues that had no apparent link with character that we assumed would nonetheless be influenced by the character manipulation. We also narrowed the number of issues from three to two in order to put further pressure on participants to give responses to the issues that were consistent with their final decision. This was achieved by removing the issue of whether or not Tom was 'at

work'. This left the issues of whether or not the contract could be interpreted to include being drunk within the meaning of gross misconduct and whether the previous case where an employee was not disciplined despite being drunk was similar to the allegations in the present case. In addition, there was only a single character manipulation at the outset and not a second character manipulation at the conclusion of the evidence. Correspondingly, participants were only asked to give a final decision and were not asked to give a preliminary indication. To further try to ensure that character was not formally an issue, participants were this time told that Tom had admitted to being drunk and disruptive and had been dismissed for gross misconduct. Their task was explained to be as an independent arbitrator considering Tom's appeal against gross misconduct on the two narrow formal grounds described above. For ease of analysis and because the previous neutral conditions in Study 3 provided relatively little additional insight, we confined the manipulations simply to a good and a bad condition.

We predicted that in these circumstances, we would see both an influence of character both on the final decision and on participants' determination of issues that bore no logical relationship with character, primarily because we considered that participants would want to favour the party they had greater sympathy for, but could not do so unless they also determined the corresponding issues in a way that would be consistent with their final decision. The R statistical analysis was preregistered with Open Science Framework.

### 4.5.1 Method

#### 4.5.1.1 Participants

One hundred participants recruited using the online survey platform Prolific completed the survey (79 females, 18 males, 3 other; aged 19 to 66, M age 33.0, SD 11.4; 23% were students; 52% in full time employment, 11% in part time employment, 37%

unemployed or other; all were UK nationals of whom 96% lived in England and 4% lived in Wales and 93% were born in the UK). A range of simulations were undertaken to understand the power of the sample size. For example, assuming standardised random intercept variances for participant and measure of 0.1 and 0.8 respectively, and a standardised residual variance of 0.5, the sample had an 80% power to detect an effect size of r2=0.19. Due to the previously mentioned information policy by Prolific, some participant data had expired by the time the analysis was undertaken and for that reason could not be used. Participants were paid £0.50 for their time. The study was conducted in accordance with approval obtained from UCL's Research Ethics Committee (EP/2018/005). Informed consent was obtained from each individual in advance of participation by providing them with study information and a consent form to agree to.

#### 4.5.1.2 Design

For Study 3 we used a between participant design with one independent variable and two levels in which all participants viewed a single set of materials but where all were subjected to a character manipulation at the outset (half seeing a good character manipulation and half seeing a bad character manipulation). This time there was no control condition.

#### 4.5.1.3 Materials

Study 4 was based on the same materials as Study 3. However, in order to eliminate the 3 issues that depended on character, the context used was a disciplinary appeal.

At the outset, participants were provided with the same confidential character references as with Study 3, but a second character reference was not provided.

Participants were advised that the employee Tom Clarke admitted to the investigating officer that he was intoxicated and disruptive at the awards ceremony, and that he had previously been dismissed for gross misconduct.

Participants were asked to determine only two issues: (1) whether the ambiguous clause of the employment contract classed intoxication as gross misconduct; and (2) whether the previous case was analogous to Tom's case such that it would be inconsistent to dismiss him.

The text of the contract and the previous case were provided to participants in the same form as in Study 3.

As before, participants were provided with arguments from both parties.

#### 4.5.1.4 Measures

For Survey 4, participants were only asked to give responses after all the evidence was complete. They were asked which party their final decision was in favour of, Tom Clarke or Paragon, and the presentation of these two options was randomised by the Qualtrics survey platform as before. They were then asked how confident they were in that decision, as before using a Likert scale from 0 to 5 with a verbal equivalent scale running from the left to right hand extremes as previously used in Study 3.

Participants were then asked to indicate their agreement to three issues. These were (1) that Tom was on probation (the attention check), (2) that being intoxicated was not gross misconduct under Paragon's disciplinary policy, so Tom should not have been dismissed, and (3) the previous incident were a female member of staff was disciplined after drinking champagne was similar to the disciplinary charge against Tom, so it was unfair to dismiss Tom. For each, participants again indicated their agreement on the same scale from -5 via 0 to

+5 with verbal equivalents on a scale running from the left hand extreme to the right hand as used in Study 3.

#### 4.5.1.5 Procedure

Again, participants were recruited online and participated in the survey in a place of their choosing, using their own device. On referral from the Prolific platform, participants were provided with the study information form. They were next asked to complete the consent form comprising a number of statements to which an affirmative answer was required in order to participate in the survey. Their anonymous user identification was collected to enable subsequent matching of demographic data without compromising the participants' anonymity.

Participants were told that they were being asked to play the role of an independent arbitrator who had been asked to decide a workplace disciplinary appeal. They were told that they should make their decision on the basis of the information and arguments from both sides, that they were not expected to have legal knowledge, and that they should use common sense when making their decision.

Before viewing the evidence and arguments from both sides, participants were presented with the same character references provided at the start of the evidence as before. Given the two conditions, half of participants were provided with the good character reference for Tom and half were provided with the bad character reference. Randomisation was undertaken by the survey platform. Unlike with Experiment 3, there was not a second character manipulation later in the survey.

Participants were again told that Tom was a full-time employee of Paragon, having completed his probation some time before. They were similarly told that at an awards presentation at a hotel, there was a free bar which some employees took advantage of,

including Tom and that during the awards presentation, he was loud and boorish, shouting abuse in a slurred voice, and was so drunk he could hardly stand. Participants were told that Tom had admitted being intoxicated and disruptive to the investigating officer and had subsequently been dismissed for gross misconduct following an internal disciplinary hearing.

Participants were advised that the appeal against dismissal was on two grounds. The first ground was that under Paragon's disciplinary policy being drunk or disorderly was not classed as gross misconduct and therefore the company was not authorised to dismiss Tom on that basis. The second ground was that even if being drunk and disorderly amounted to gross misconduct, Paragon had not previously dismissed employees for that behaviour, so it would be unfair due to inconsistency to dismiss Tom.

Next participants saw the relevant ambiguous extracts of the disciplinary policy previously provided in Study 3 together with details of the previous precedent in which a female employee was described as becoming unwell after an impromptu celebration of a sales target at which several bottles of champagne were served and where no disciplinary proceedings were instituted.

At the conclusion of the evidence, participants read the arguments put forward by the representatives of the parties, reiterating their relevant positions. Finally, participants were asked to indicate their responses on the measures previously described.

### 4.5.2 Results

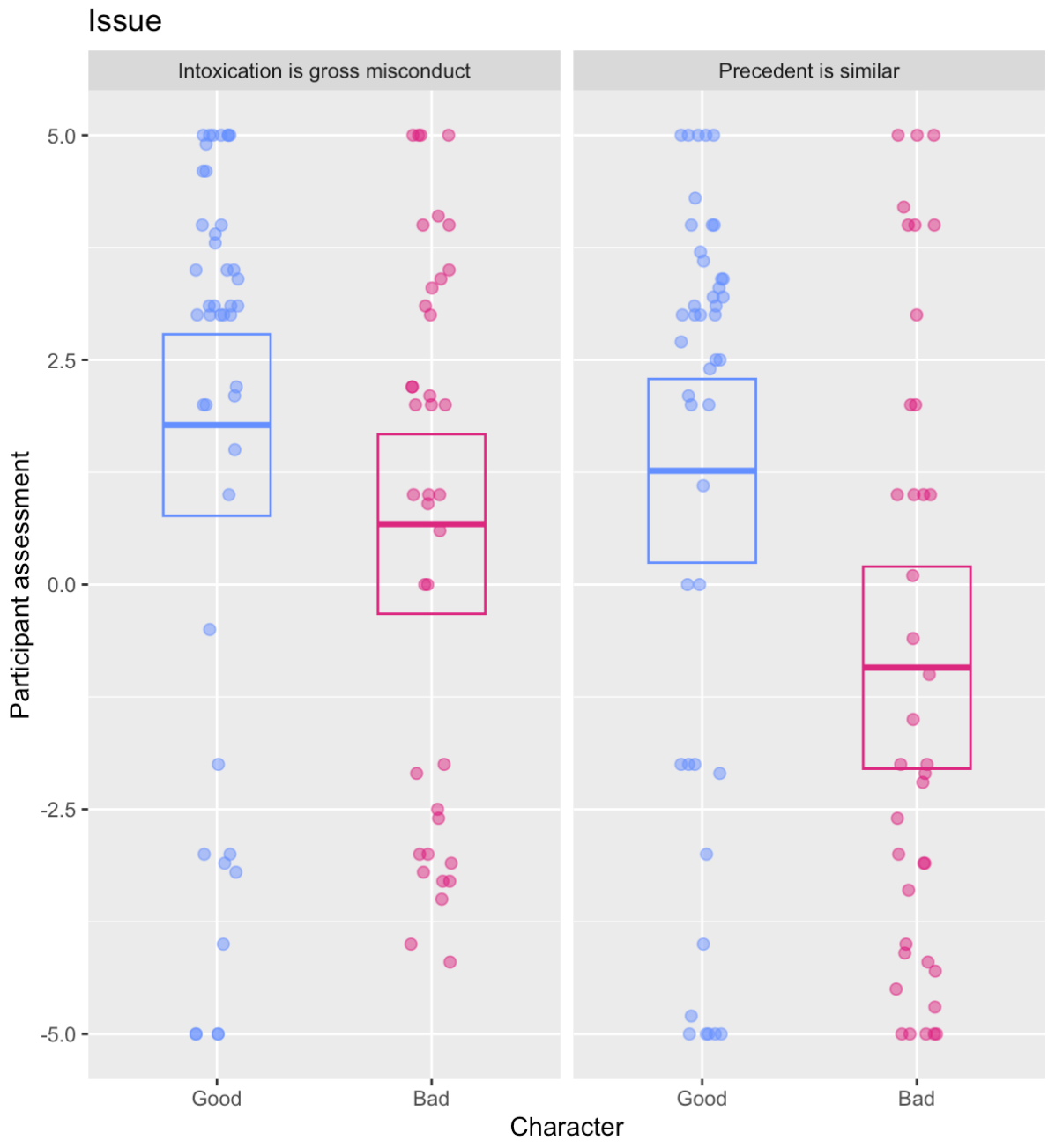
Of the 101 participants who completed the survey, 19 (19%) failed the attention check question by incorrectly indicating that Tom was an employee on probation. This data was excluded from the analysis.

As with previous studies, the character manipulation had a significant effect on final

decisions. Those participants in the good character condition were much more likely to find in favour of Tom compared to Paragon (36:8) compared to those participants in the bad character condition (16:22). Again, a Fisher's Exact Test confirmed that this pattern was statistically significant ($p < 0.001$, OR = 0.17, 95% CI [0, 0.42]). Again, participants were very confident in their decisions (M=3.5, SD=0.84) with no statistically significant difference between the conditions (B=0.16, t=0.83, p=0.41).

In relation to the extralegal effect of character on issues not ostensibly related to character, there was a difference between responses depending on the condition that the participants were in, suggesting that in accordance with our predictions, there was an extralegal effect of character consistent with Studies 1 and 2. In the good character condition, mean responses were 1.52, whereas in the bad condition, mean responses were -0.13, a difference of 1.65, see Figure 7. A mixed effect model with random effects for participant and issue confirmed that this difference was statistically significant (B=1.13 (95% CI [0.32, 1.95]), df=80, t=2.73, p=0.008). Mean responses for the issue of whether the previous precedent was similar to the present allegations were 1.27 and -0.92 respectively, a difference of 2.19 (95% CI [0.70, 3.68]). Mean responses for the issue of whether the behaviour amounted to gross misconduct according to the contract were 1.78 and 0.67 respectively, a difference of 1.11 (95% CI [2.51, -0.31]). The issue of similarity was significant in isolation (t(80)=2.91, p=0.005, Cohen's D=0.65) whereas the gross misconduct issue was marginally significant in isolation (t(80)=1.56, p=0.12, Cohen's D=0.34).

*Figure 7. Participant assessment of issues by character manipulation showing mean, 95% confidence interval, and data points for Study 4.*



### 4.5.3 Discussion

The findings from Study 4 suggest that a requirement to appear consistent is a necessary ingredient for some of the extralegal effects we have been studying. In all studies character appeared to have a significant influence on final decisions, even those where character was legally irrelevant. In Studies 1, 2, and 4, neither the final decision nor the issues that fed into that final decision ought legally to have been influenced by character, but it appeared that many participants nonetheless took character into account in a legally impermissible way. Given these findings, a similar pattern regarding decisions probably manifested itself in Study 3, but it would have been difficult to isolate this because in that study character was at least partially legally relevant to the final decision.

More noteworthy was the apparent extralegal influence of character on the issues that fed into those final decisions. This was shown in Studies 1, 2, and 4, but not Study 3. The difference between Studies 3 and 4 seems to be explained primarily by the pressure or absence of pressure on participants to determine the issues consistently with their final decisions. In Study 3, although the same measures were used, there was less pressure on participants to give a response to those issues in a way that was consistent with their final decision because the issues that fed into that decision included both issues that would be expected to be related to character as well as those that would not. In other words, even if their final decision might have been influenced by character, this could reasonably be justified by their responses to issues that would be expected to be influenced by character. There was seemingly little or no need to also determine the issues unrelated to character consistently with the final decision. However, asking participants to determine those same issues unrelated to character in the much more constrained environment of Study 4 did show a significant effect. In Study 4 the final decision was much more closely linked to their findings on issues unrelated to character. Those participants whose final decisions were influenced by character would also have needed to determine the issues feeding into those decisions consistently with those final decisions to appear consistent. This is exactly what we saw: both final decisions and decisions on issues legally unrelated to character were

nonetheless influenced by character.

These findings seem to support our theory that these extralegal effects are better explained by prudential behaviour by adjudicators rather than irrationality or cognitive limitations. This is because there does not seem to be the expected relationship between task complexity and these extralegal effects as might be predicted by irrationality type explanations. Rather, these extralegal patterns seem to manifest themselves in the simpler contexts of Studies 1, 2, and 4, but not the more complex context of Study 3. Our explanation would be that the key difference of Study 3 is that the adjudicatory task included issues to which character was relevant such that character could legitimately be taken into account. There was thus much less pressure on participants to bend the rules.

Studies 1 to 4 still left open the question raised in Study 3 as to how participants would respond to a dual manipulation where the issues were similarly constrained as per that study. Our prediction was that preliminary indications would show a similar pattern to Study 4 because they would effectively be a single manipulation. More uncertain was how participants would behave at the final decision. Two possibilities seemed open: that participants would switch their assessments of the issues influenced by character on the basis that the preliminary indication was described as confidential to them, or that participants would stick with their initial assessment because they would want to appear consistent to the observers collecting the data.

## 4.6 STUDY 5

For Study 5 we were interested in returning to examine the effect of the dual character manipulation on final decisions posed in Study 3, but this time in a context where the adjudicatory environment was sufficiently constrained (as with Studies 1, 2, and 4). We were aware from these previous experiments that this led to the extralegal effect of character where there was a single character manipulation, but the effect of a second character manipulation

still remained open. For a single character manipulation at the preliminary indication stage, our prediction would have been that we would have seen the extralegal influence of character, as this would have been a very similar paradigm to our earlier research. However, the effect of a dual character manipulation prior to a final decision was still uncertain. Our original prediction was that we would see extralegal effects of character in the opposite direction, but this assumes that there would be limited effect of the participant previously disclosing their preliminary indication to observers. We noted that while participants had always been instructed that their preliminary indication was confidential to them and would not be seen by the ostensible parties to the adjudication, obviously their preliminary indication would be visible to the experimenters and this might influence them to be consistent. Thus a second possibility would be that for their final decision, participants would instead stick with their preliminary indication, in order to appear consistent to the observers, the experimenters.

In order to eliminate this potential effect of participants making their final decision appear consistent with their preliminary decision to the eyes of an observer, we decided next to conduct an experiment where this potential effect was minimised. Thus, for Study 5, we provided a dual character manipulation, but did not ask the participants to disclosure their preliminary indication following the first character manipulation. Our prediction was that we would see an effect of character at the final decision stage (in line with the second character manipulation) because there would be no risk of the participant being perceived as potentially inconsistent. Study 5 therefore used the constrained context of Study 4 where participants were again only asked to give a final decision. The only material difference was that this time there were two character manipulations as originally envisaged for this paradigm, one at the outset before the evidence and one after the evidence and prior to the final decision. The R statistical analysis was preregistered with Open Science Framework.

### 4.6.1 Method

#### 4.6.1.1 Participants

One hundred participants recruited using the online survey platform Prolific completed the survey (75 females, 22 males, 3 other; aged 31 to 60, M age 31.3, SD 10.0; 21% were students; 62% in full time employment, 19% in part time employment, 19% unemployed or other; all were UK nationals of whom 93% lived in England and 7% lived in Wales and 96% were born in the UK). A range of simulations were undertaken to understand the power of the sample size. As with the previous experiment, assuming standardised random intercept variances for participant and measure of 0.1 and 0.8 respectively, and a standardised residual variance of 0.5, the sample had an 80% power to detect an effect size of r2=0.19. Due to the previously mentioned information policy by Prolific, some participant data had expired by the time the analysis was undertaken and for that reason could not be used. Participants were paid £0.50 for their time. The study was conducted in accordance with approval obtained from UCL's Research Ethics Committee (EP/2018/005). Informed consent was obtained from each individual in advance of participation by providing them with study information and a consent form to agree to.

#### 4.6.1.2 Design

Study 5 used a between participant design with one independent variable and two levels in which all participants viewed a single set of materials but where the two groups were subjected to two character manipulations, either good then bad, or bad then good, before giving their final decision. As with Study 4, there was no control condition.

#### 4.6.1.3 Materials

Study 5 used the same materials as Study 4, being an appeal based on agreed facts, with the exception of the two issues of the interpretation of the contractual clause and the question of whether the previous precedent was similar to the agreed facts in the appeal. However, because of the planned dual character manipulation, Study 5 also used a second, opposing, character manipulation. This second character manipulation was the same as used in Study 3.

#### 4.6.1.4 Measures

Study 5 used the same measures as Study 4.

#### 4.6.1.5 Procedure

Study 5 followed a very similar procedure to Study 4. Again participants were told that they were conducting a workplace disciplinary appeal. However, due to the dual character manipulation, participants were told at the outset that there may be some further information becoming available and that they would consider the available information and arguments but not reach a final decision until they saw any further information. Participants were told that once they had receive the further information, they would make a final decision.

As before, participants were provided with a confidential character reference at the outset that indicated that Tom was either of good or bad character, depending on which condition they had been assigned to. They were then presented with the same information and arguments as per Study 4. However, following the arguments, they were given interim instructions that they were waiting for further information and not to take a final decision

until they had seen the further information that had just become available. Participants were then provided with the second character reference, presented as per Experiment 3. The nature of the character reference was the inverse of that provided at the initial stage and the explanation given for this was the same as with Experiment 3, that either Tom or his supervisor had blackmailed the administrator to substitute a more favourable or less favourable reference.

Following presentation of the second reference, participants were asked to indicate their views on the same measures used in Study 4, namely which party their final decision was in favour of, how confident they were in that decision, agreement with the three issues (the probation attention check, whether being intoxicated amounted to gross misconduct, and whether the previous incident was similar to the present charges). As before, participants were able to refer back to the information and arguments of the parties.

### 4.6.2 Results

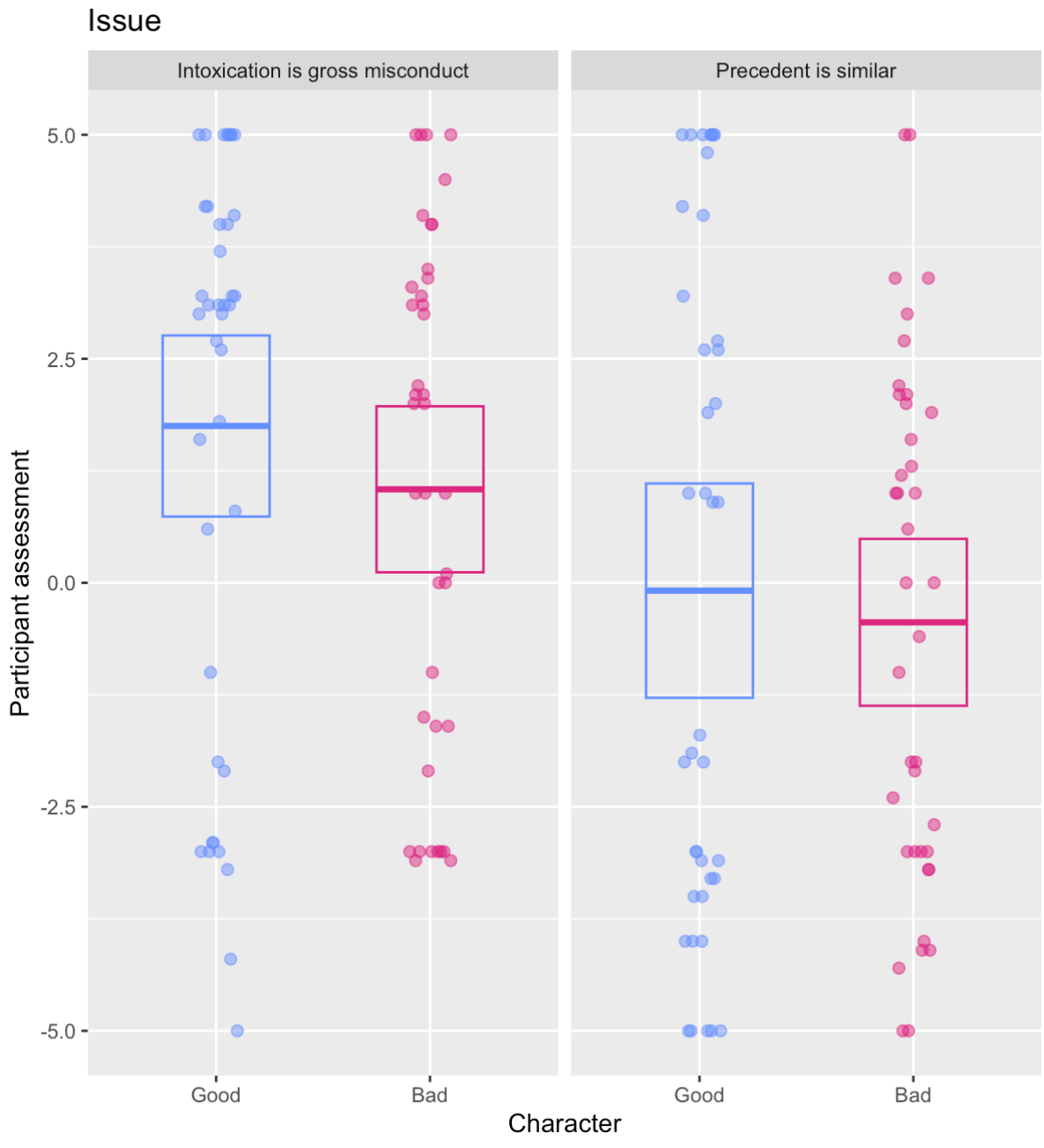
100 participants completed the survey, of which 21 (21%) failed the attention check, a very similar proportion as with Studies 3 and 4.

As with Studies 3 and 4, the character manipulation had a significant effect on final decisions. Those participants in the good character condition were much more likely to find in favour of Tom compared to Paragon (27:13) compared to those participants in the bad character condition (13:26). A Fisher's Exact Test showed that this pattern was statistically significant ($p = 0.002$, OR = 0.25, 95% CI [0, 0.59]) though the effect size was smaller than for Study 4. Again, participants appeared very confident of their decisions with a mean response of 3.78 (SD 0.72) on the scale of 0-5. As with previous experiments, there was no significant difference in confidence between conditions ($B=0.10$, $t=0.63$, $p=0.53$).

However, the picture in relation to the issues was not as expected. See Figure 8. This

time, there was a much less distinct effect of character on participants' responses to the issues. Although the trend was similar to Experiment 4, the effect was considerably smaller. The overall average response to the issues was 0.83 in the good character condition and 0.30 in the bad character condition, a difference of only 0.53, a much smaller difference than predicted, and a much smaller difference than seen in Study 4. A mixed effects model with random effects for participant and issue indicated that the difference did not reach statistical significance on a standard 0.05 alpha level: (B=0.31 (95% CI [-0.51, 1.13]), df=77, t=0.74, p=0.46).

*Figure 8. Participant assessment of issues following dual character manipulation showing mean, 95% confidence interval, and data points for Study 5.*




Correspondingly, the differences between conditions on the individual issues did not reach statistical significance either. Mean responses for the issue of whether the behaviour

amounted to gross misconduct according to the contract were 1.04 and 1.75, a difference of 0.71 (95% CI [-0.64, 2.06]). Mean responses for the issue of whether the previous precedent was similar to the present allegations was -0.44 and -0.09, a difference of 0.35 (95% CI [-1.14, 1.85]). The issue of gross misconduct (t(77)=1.04, p=0.30, Cohen's D=0.23) and neither was the issue of similarity was not significant in isolation (t(77)=0.47, p=0.64, Cohen's D=0.11).

### 4.6.3 Discussion

The findings of Study 5 were somewhat unexpected. It seemed that adding a second (and opposite) character manipulation mid-way through the experiment had the effect of moderating much of the extralegal effects on responses to the issues. Study 5 contrasts with Study 4 where a single character manipulation at the outset had quite a distinct extralegal effect on issues. This is particularly interesting if we recall one of the incidental findings from Study 3 that suggested that the second character manipulation appeared to be a stronger manipulation than the first character manipulation. However, notwithstanding the apparent strength of the second character manipulation, it did not appear to have a significant effect in the overall context of Study 5. Given that the only difference between Studies 4 and 5 was the introduction of a second character manipulation, and we know from Studies 1, 2, and 4 that character manipulations generally lead to extra-legal effects, the use of two manipulations appeared to have some kind of moderating effect.

One possible explanation for the moderating effect of the two manipulations is the existence of counterfactual circumstances. Certain counterfactual circumstances allow an observer to glean more about the decision maker's thought process and thereby risk unveiling that the decision maker is impermissibly taking extralegal information into account. For example, if a decision maker is asked to make a decision on two otherwise identical cases that differ by only one (legally irrelevant) dimension, if the decision maker makes different decisions across the two cases, then it is trivial for an observer to infer that the decision

maker has impermissibly taken legally irrelevant information into account. Thus, in concrete cases, the existence of concrete counterfactuals has been empirically shown to be a factor that seems to moderate these sorts of extralegal effects. For example, where extralegal effects exist in individual cases, presenting two cases side-by-side that differ only by the counterfactual is sufficient to eliminate the effect (Nadler, 2012, p. 26; Sood & Darley, 2012, pp. 1343–1344). This could be effective because the existence of the counterfactuals is enough to show on an individual basis that the participant has impermissibly taken extralegal information into account and there is thus little or no ambiguity. However, Study 5 is distinct from previous research because in the context of our study, it would not in fact be possible for an observer to glean information about the individual decision maker, because each decision maker only made a single decision. However, it is possible that the dual manipulation alerted the decision makers to the risk of counterfactuals, thus moderating their behaviour and making them more cautious.

## 4.7 STUDY 6

For Study 6 we looked to replicate some of our previous findings, while also extending the research to examine the effect of asking participants to give two indications of their views: a preliminary indication and a final decision. We did this by asking one group of participants to only give a final decision after experiencing two opposing character manipulations, one prior to considering the materials, and the other after considering the materials. This was essentially a replication of the previous experiment, Experiment 5. The other group of participants were asked to give both a preliminary indication after experiencing the first character manipulation, followed by a final decision after experiencing the second, opposite, character manipulation. Considering their preliminary indications in response to a single character manipulation was essentially a replication of our previous experiments with a single character manipulation, namely Experiments 1, 2, and 4, albeit at an interim stage in the process rather than at a final stage in the process. However, asking this group to give both a preliminary indication and a final decision in a constrained paradigm

allowed us to test the novel hypothesis of whether we could invoke the types of extralegal effects using dual manipulations, the question we had initially raised in Experiment 3.

In terms of our hypotheses, for the participants subjected to a dual character manipulation before being asked to give only a final decision, this was essentially a replication of Experiment 5, so we expected to see similar results. Thus, while some evidence of an effect of character in accordance with the second character manipulation was expected, this was predicted to be much smaller than the effects seen in our experiments with only a single character manipulation, namely Experiments 1, 2, and 4.

In relation to participants' responses to preliminary indications, this was very similar to our previous experiments involving a single manipulation, namely Experiments 1, 2, and 4. While Experiment 6 was not completely identical to those experiments because it involved a preliminary indication rather than a final decision, we expected to see a similar pattern, with extralegal effects of the single character manipulation on participants' responses.

The novel dimension to Experiment 6 was that we tested a dual manipulation in a constrained scenario, rather than an unconstrained scenario as tested in Experiment 3. Thus, for those participants who had sympathy or antipathy to the employee or the employer, the constrained paradigm meant that they could only favour one party or the other by impermissibly taking character into account in relation to issues that bore no relation to character. Assuming that participants who gave a preliminary indication demonstrated extralegal effects of the first character manipulation, this raised a question as to how they would then respond to the second character manipulation. One possibility was that they would simply take character into account in the opposite direction. We had advised participants that their preliminary indication was confidential to them and would not be seen by the parties. However, it was obvious that the preliminary indication would be available to the experimenters. Thus, if participants impermissibly took character into account for the preliminary indication, then took it into account in the opposite direction for the final decision, it would have been transparent from the incoherence of their two responses (ie that

the being drunk was and was not gross misconduct according to the contract and that the previous precedent was both similar and dissimilar from the circumstances that they had to consider) that they were impermissibly taking character into account. We therefore predicted that, provided there was an extralegal effect of character after the first character manipulation, participants would not change their assessments of the issues at the final decision so as to avoid looking inconsistent to the experimenters.

## 4.7.1 Method

### 4.7.1.1 Participants

Two hundred participants recruited using the online survey platform Prolific completed the survey (143 females, 57 males; aged 18 to 64, M age 42.4, SD 11.8; 8% were students; 56% in full time employment, 19% in part time employment, 25% unemployed or other; all were UK nationals of whom 96% lived in England and 4% lived in Wales and 94% were born in the UK). The comparisons of interest were essentially the same as Experiments 4 and 5, thus the previous simulations prepared for those experiments were similarly applicable. As before, assuming standardised random intercept variances for participant and measure of 0.1 and 0.8 respectively, and a standardised residual variance of 0.5, the sample had an 80% power to detect an effect size of r2=0.19. Participants were paid £0.75 for their time. The study was conducted in accordance with approval obtained from UCL's Research Ethics Committee (EP/2018/005). Informed consent was obtained from each individual in advance of participation by providing them with study information and a consent form to agree to.

### 4.7.1.2 Design

Study 6 used a 2 × 2 between participant design in which all participants viewed a single set of materials but where half of participants gave both a preliminary indication and a final decision and where half gave only a final decision. As with Studies 3 and 5, all participants were subjected to two character manipulations, either good then bad, or bad then good. There was no control condition.

#### 4.7.1.3 Materials

The materials used in Study 6 were the same as Studies 4 and 5.

#### 4.7.1.4 Measures

Study 6 used the same measures as Studies 4 and 5, namely a preliminary indication and/or a final decision in favour of one party or the other and an preliminary and/or final assessment of the two issues in the appeal, namely agreement on a -5 to +5 point Likert scale. The same scale was used for the attention check regarding the issue of probation, to confirm that participants had paid attention to the materials. In addition, for external validity and in order to see if any further insight could be gleaned into participants' thinking, participants were also asked to give reasons for their preliminary indication and decision, as with Studies 1 and 2.

#### 4.7.1.5 Procedure

The procedure adopted in Study 6 followed a similar procedure to previous experiments. All participants saw an initial character manipulation prior to the materials. This character manipulation was either good or bad, with half of participants randomly assigned to

the good condition or the bad. Participants were then shown the same materials as with Studies 4 and 5 but told that further information may be becoming available. Thereafter, half of the participants were asked to give a preliminary indication of their decision and assessment of the issues on the basis that this was confidential to them and would not be seen by the parties. Next, all participants were shown a second character manipulation, and this was the opposite of what was seen at first: ie, those shown a good character manipulation at the outset were then shown a bad character manipulation and vice versa.

As a result, a quarter of participants were shown a good character reference followed by a bad character reference before making a single final decision. A quarter of participants were shown a bad character reference followed by a good character reference before making a single final decision. A quarter of participants were shown a good character reference before being asked to make a preliminary assessment, and were then shown a bad character reference before being asked to make a final decision. The final quarter were shown a bad character reference before being asked to make a preliminary assessment, followed by a good character reference and being asked to make a final decision.

### 4.7.2 Results

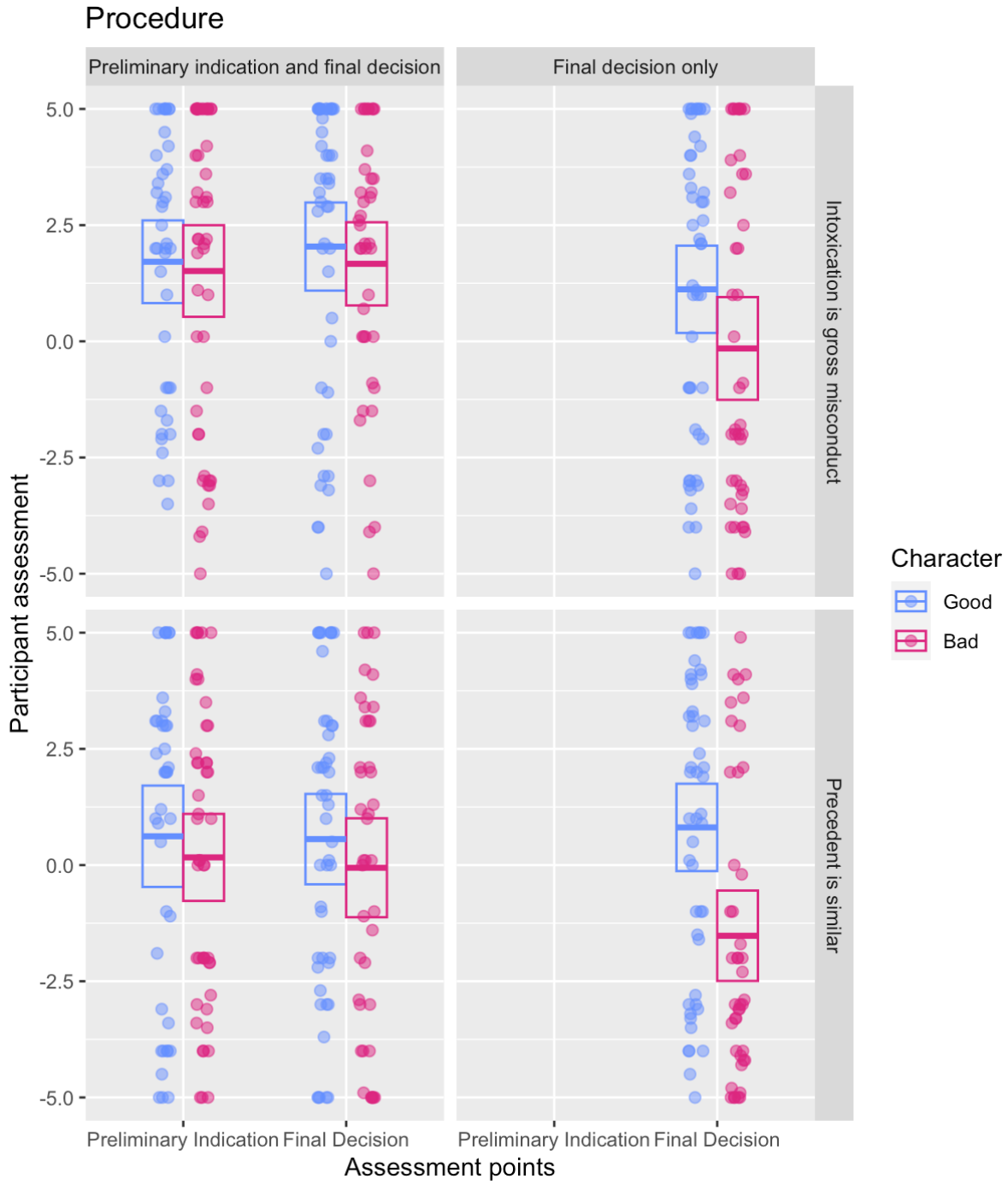
200 participants completed the survey, of which 24 (12%) failed the attention check, a slightly lower proportion than with Studies 3, 4, and 5.

For those participants giving only a final decision, consistent with all previous studies, there was a significant effect of character on actual decisions. In the good character condition, participants were much more likely to find in favour of Tom (33:14) than in the bad character condition (11:32). A Fisher's Exact Test confirmed this was significant ($p < 0.001$, OR=6.69, 95% CI [2.48, 19.34]). A similar pattern was seen in responses to the issues, see Figure 9, right side panels. A mixed effects model with random effects for participant and issue confirmed that the difference reached statistical significance ($p<0.01$, B=1.80 (95% CI

[0.55, 3.05]), df=87, t=2.84).

However, in relation to the participants giving a preliminary indication, a more interesting pattern was seen. Unlike previous studies, there was very little difference between those finding in favour of Tom in the good character condition (28:12) compared to in the bad character condition (27:19). Using a Fisher's Exact Test, this was unsurprisingly not significant (p=0.37, OR= 1.63, 95% CI [0.61, 4.46]). Correspondingly, there was also very little difference in responses to the issues, see Figure 9, left side panels. A mixed effects model with random effects for participant and issue did not reach statistical significance (p<0.61, B=0.33 (95% CI [-0.93, 1.59]), df=84, t=0.51). Although there was a marginally more sympathetic response to the employee from participants in the good character condition, the difference was marginal, with practically no difference between the groups.

*Figure 9. Participant assessment on issues by character for participants giving both a preliminary indication and a final decision and for participants giving only a final decision in Study 6.*



This interesting pattern was reflected in the responses when these same participants gave their final decisions after seeing the second, opposing, manipulation. Responses to the final decision were more akin to those seen in previous experiments, with evidence of participants favouring Tom more in the good character condition (33:13) than in the bad character condition (19:21). This pattern was statistically significant (p=0.03, OR=2.77, 95% CI [1.05, 7.56]). However, as with the measures taken at the preliminary stage, there was very little difference between the responses to the issues, with practically no movement after the second character manipulation, see Figure 9, left hand panels. A mixed effects model confirmed this (p=0.43, B=0.49 (95% CI [-0.73, 1.72]), df=84, t=0.80). To some extent this was consistent with our predictions of a lack of change in participants' responses to the issues between the preliminary indications and final decisions, albeit from a different starting point in that we had expected there to be a difference between responses at the preliminary indication stage rather than no difference.

In terms of confidence, all participants expressed relatively high confidence in their decisions on the 5 point scale, though, as one might expect, with a slightly lower confidence for preliminary indication (3.26) than for final decisions without a previous preliminary indication (3.72) or final decisions following a preliminary indication (3.70). There was no statistically significant difference in confidence between those in the good character conditions compared to the bad character conditions.

### 4.7.3 Discussion

There were a number of interesting statistical findings from Study 6. Firstly, in relation to the replication of Study 5 with a dual character manipulation where participants gave a final decision only, this time the result was significant. After a dual character manipulation, participants were much more likely to find in favour of Tom where the second manipulation was that he was of good character and against him when he was of bad character. More pertinently, participants also determined the issues, that ought not to have

been influenced by character, in accordance with the second character manipulation that they were subjected to, thereby demonstrating a clear extralegal effect. This was a little unexpected, given that the effect in the similarly powered Study 5 had not reached statistical significance, even if the effects in that study had been in the same direction. The only difference between these studies was the addition of a free text response field for participants to explain their decision, but this seemed unlikely to have aggravated the extralegal effects given that previous research has suggested that giving reasons tends instead to moderate extralegal effects ((Liu, 2018; Tetlock, 1983; Wistrich et al., 2004, p. 1324)). Given the materials were also almost identical, the pattern seen in either Study 5 or this aspect of Study 6 may have been due to stochastic factors. This uncertainty suggested further scrutiny was justified.

A second interesting finding was the apparent elimination of extralegal effects at the preliminary indication stage. This was in stark contrast to the extralegal effects seen where participants were asked to give a final decision only, and also in contrast to the extralegal effects seen in previous studies, in particular Studies 1, 2, and 4. The differences that stood out from previous studies were that (1) the indication was preliminary rather than final, and (2) participants were asked to give reasons for their preliminary indications. Here, the moderating effect of giving reasons on the extralegal effects seems more plausible given that previous research has suggested that it has this effect ((Liu, 2018; Tetlock, 1983; Wistrich et al., 2004, p. 1324)). It may have been the case that the inchoate nature of the materials, combined with a requirement to justify their thinking made participants much more cautious about impermissibly taking character into account in decisions that are supposed to have nothing to do with character. If such an effect is confirmed, it could prove quite a powerful tool in addressing extralegal effects where these are unwanted.

The third interesting finding was an apparent consistency effect for participants giving a preliminary indication and then a final decision. This was not quite as expected, because we had presumed that we would see extralegal effects at the preliminary indication stage, which participants would then be consistent with at the final decision stage. Notwithstanding the

lack of extralegal effects at the preliminary stage, we did see an apparent consistency effect at the final decision stage whereby participants who did not demonstrate extralegal influences at the preliminary indication stage did not demonstrate them at the final decision stage either. This was probably evidence of a consistency effect given that, by contrast, we saw that participants who only gave a final decision showed very clear evidence of extralegal influences.

Given the apparently anomalous findings, in particular the moderation of extralegal effects at the preliminary indication stage, and the fact that we had introduced a seemingly innocuous requirement go give reasons, we considered it appropriate to rerun the experiment without the requirement to provide reasons, and also to replicate the dual character manipulation in order to see if the extralegal effect could be repeated, contrary to the findings in Study 5.

## 4.8 STUDY 7

Study 7 was identical to Study 6, save for the removal of the requirement for participants to give reasons in a text box when they gave a preliminary indication or a final decision. In Study 6, we had seen no extralegal effects of character and it seemed likely that the requirement to give reasons had disrupted the extralegal effect as had been suggested in previous research. We predicted that without the requirement to give reasons, we would see an extralegal effect as originally predicted in Study 6. Equally, provided an extralegal effect of character was established at the preliminary indication stage, we predicted that participants would then seek to remain consistent with this at the final decision stage, notwithstanding the reverse character manipulation. This was predicted to happen even though the effect would be to disadvantage the party that they would be likely to have more sympathy with.

We also sought to replicate the effect seen in Study 6 where a dual character manipulation with no preliminary indication but a final decision did see extralegal effects in

the direction of the second character manipulation. We predicted that these would manifest themselves notwithstanding the removal of the requirement to give reasons, given that the requirement to give reasons seems generally to moderate extralegal effects.

## 4.8.1 Method

### 4.8.1.1 Participants

Two hundred participants recruited using the online survey platform Prolific completed the survey (134 females, 66 males; aged 19 to 65, M age 38.5, SD 10.9; 13% were students; 51% in full time employment, 28% in part time employment, 21% unemployed or other; all were UK nationals of whom 95% lived in England and 5% lived in Wales and 95% were born in the UK). Given the underlying comparisons were equivalent to the previous experiments, the previous simulations prepared for those experiments were equally applicable. Assuming standardised random intercept variances for participant and measure of 0.1 and 0.8 respectively, and a standardised residual variance of 0.5, the sample had an 80% power to detect an effect size of r2=0.19. Participants were paid £0.60 for their time. The study was conducted in accordance with approval obtained from UCL's Research Ethics Committee (EP/2018/005). Informed consent was obtained from each individual in advance of participation by providing them with study information and a consent form to agree to.

### 4.8.1.2 Design

As with Study 6, Study 7 used a 2 × 2 between participant design in which all participants viewed a single set of materials but where half of participants gave both a preliminary indication and a final decision and where half gave only a final decision. As with Studies 3, 5, and 6, all participants were subjected to two character manipulations, either

good then bad, or bad then good. There was no control condition.

#### 4.8.1.3 Materials

The materials used in Study 7 were the same as Study 6, with the exception of the absence of the expectation on participants to give reasons for their decisions.

#### 4.8.1.4 Measures

As discussed, Study 7 used identical measures to Study 6, other than the requirement to give reasons at the preliminary indication and final decision stages.

#### 4.8.1.5 Procedure

The procedure for Study 7 was the same as Study 6, other than the fact that participants were not required to give reasons for their preliminary indications or final decisions after they had determined the issues.

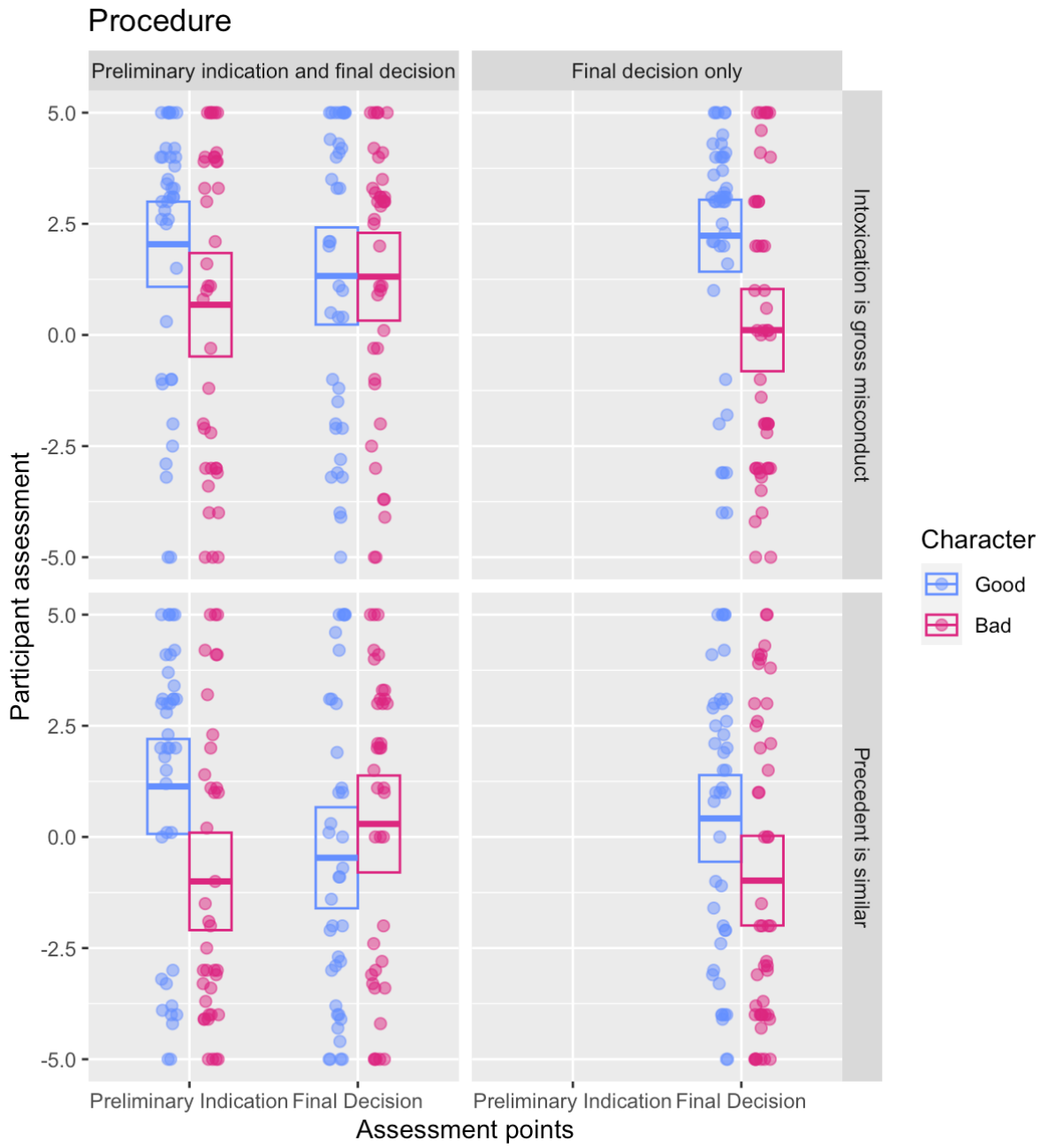
### 4.8.2 Results

200 participants completed the survey, of which 33 (16.5%) failed the attention check, a slightly higher proportion than in Study 6, but slightly lower than in Studies 3, 4, and 5.

For those participants who gave a final decision but not a preliminary indication after seeing the dual character manipulation, as predicted, this replicated Study 6, notwithstanding the removal of the requirement to give reasons for their decision. There was a significant

effect of character, such that participants were much more likely to determine the decision in accordance with the second character manipulation. In the good character condition, participants were much more likely to find in favour of Tom (33:10) than in the bad character condition (13:34). Undertaking a Fisher's Exact Test confirmed that this was significant ($p < 0.001$, OR=8.39, 95% CI [3.04, 25.17]). Correspondingly, participants' responses to the issues were influenced by the second character manipulation, see Figure 10, right hand panels. According to a mixed effects model with random effects for participant and issue, this difference was significant ($p<0.01$, B=1.76 (95% CI [0.61, 2.92], df=88, t=3.00).

*Figure 10. Participant assessment on issues by character for participants giving both a preliminary indication and a final decision and for participants giving only a final decision where participants were not asked to give reasons in Study 7.*




In relation to those participants giving a preliminary indication after a single character

manipulation, this was as predicted in Studies 6 and 7 in that there was a significant effect of character. Participants were more likely to determine the appeal in Tom's favour when he was of good character (28:11) than when he was of bad character (17:21). A Fisher's Exact Test confirmed this ($p=0.02$, $OR=3.10$, 95% CI [1.10, 9.07]). In accordance with this, participants were also likely to determine the issues in a similar way. A mixed effects model with random effects for participant and issue confirmed that this was significant ($p=0.01$, $B=1.75$ (95% CI [0.37, 3.12]), $df=75$, $t=2.50$).

For those participants who gave a final decision after a preliminary indication, the results were also as predicted in Studies 6 and 7. Participants' final decisions were, as usual, in accordance with the second character manipulation. Where the second character manipulation was good, participants tended to find in Tom's favour (27:11), whereas where the second character manipulation was bad, participants tended to find against Tom (15:24). This was significant pursuant to a Fisher's Exact Test ($p<0.01$, $OR=3.83$, 95% CI [1.38, 11.41]). More interestingly, participants final views on the issues seemed to be very much influenced by how they had determined the issues at the preliminary indication stage. Thus, because participants views on the issues at the preliminary indication stage had been quite strongly influenced by character, they seemed to maintain a similar view on the issues at the final decision stage, notwithstanding that the character manipulation was now in the opposite direction. Thus, if participants were seeking to determine the issues at the preliminary indication stage in a way that favoured the party that they had most sympathy with, they appeared to maintain this view at the final decision stage to some extent, even where this would have had the effect of favouring a side which they would then not had sympathy with. Thus there was a slight negative correlation between good character and issue of $B=-0.37$, though this was not significant ($p=0.60$, $B=-0.37$ (95% CI [-1.76, 01.01]), $df= 75$, $t=-0.529$).

In terms of confidence in their decisions, all participants expressed relatively high confidence in their decisions on the 5-point Likert scale, though, as before, participants were slightly less confident at the preliminary stage ($M=3.40$, $SD=0.80$) than at the final decision stage ($M=3.77$, $SD=0.93$ for participants not giving a preliminary indication, $M=3.76$, SD

=0.95 for this giving a preliminary indication). There were no significant differences by character.

### 4.8.3 Discussion

Study 7 explored the effect of a dual character manipulation in the context of a legal decision with constrained issues as per Studies 4, 5, and 6. In contrast to the Study 6, Study 7 did not ask participants to give reasons for their preliminary indications or final decisions. This design triggered some very different responses at the preliminary indication stage, which in turn demonstrated some illuminating findings at the final decision stage for those participants who had given a preliminary indication.

The first, important, but expected finding was the replication of the pattern seen in Study 6. This was that for participants only giving a final decision after a dual character manipulation (with no preliminary indication), the same extralegal effect seen in Study 6 was replicated, ie that decisions matched the character manipulation seen last. This was also consistent with the extralegal effects seen following single character manipulations in Studies 1, 2, and 4. A degree of uncertainty remains given the apparent inconsistency with Study 5, in which a dual character manipulation suggested effects consistent with an extralegal effect, but which did not reach statistical significance. Further work would be helpful to ensure that this extralegal effect following a dual character manipulation could be consistently replicated. However, from a theoretical perspective, there seems reason to expect extralegal effects in these circumstances if we believe that participants are looking to favour the party who they have most sympathy with. If participants see two character manipulations, a first one that is then superseded by a second, opposing, manipulation before all the evidence is complete, we might then expect that the second manipulation to have an effect. However, if a reason to disrupt this pattern is introduced, then the second manipulation may not work in the same way. This was the reason for the novel aspect introduced in Study 7, in particular to encourage participants to commit to a position after the first character manipulation. Without

the requirement to give reasons, we saw that the typical extralegal effects that we saw in a range of other studies were similarly triggered at the preliminary indication stage.

Once we established that there were extralegal effects at the preliminary indication stage, this enabled us to test a new hypothesis by then introducing the second character manipulation. Unlike in the condition where participants only gave a final decision, participants who had given a preliminary indication had committed to a particular view of the issues. If they then changed their view of the issues, this would look very inconsistent to any observer. For example, if they indicated at the preliminary indication stage that being intoxicated under the disciplinary policy amounted to gross misconduct, and then found at the final decision stage that it did not amount to gross misconduct, the participant would quite clearly be indicating to an observer that they impermissibly took character into account. Though participants were advised that the preliminary indication was confidential to them, and would not be seen by the putative parties, it would have been obvious that their responses would be visible to the experimenters. Thus, we predicted that many participants would stick with their initial assessments of the issues rather than switching them, even if this had the effect of disadvantaging the party that they had most sympathy with at the point of the final decision. This is exactly what we saw. Therefore it seems that in the context of these studies, participants prioritised appearing consistent over favouring the sympathetic party.

As an incidental note, the latter effect demonstrates an observer effect that experimenters should be alive to when conducting these types of experiments. It seems that reassurances given to participants that their responses will not be visible to the parties does not affect participants' sensitivity to the experimenters' awareness of their responses. This may apparently affect responses in a way that needs to be taken into account when drawing inferences about likely adjudicatory behaviour in the real world.

Finally, the stark difference between participants' preliminary indications between Studies 6 and 7 suggests a interesting potential means of discouraging extralegal effects that deserves further exploration. It appeared that the requirement to give reasons at the inchoate,

preliminary, stage of the proceedings was very effective at discouraging extralegal effects. Participants simply asked to give a preliminary indication showed distinct extralegal effects. But when participants were also asked to give reasons at the preliminary indication stage, the effect all but disappeared. This suggested some sensitivity to the impermissibility of taking impermissible factors into account which the obligation to give reasons may have highlighted in the minds of participants.

## 4.9 GENERAL DISCUSSION

### 4.9.1 Summary

This series of studies began with the aim of testing two competing theories to account for the extralegal effects that are sometimes seen in the context of legal adjudication. In one corner we had the current defending champion, the theory that these extralegal effects are evidence of irrationality or bias on the part of adjudicators, caused due to either or both of the cognitive challenge or the complexity of legal decision making. In the opposite corner was the developing theory that these effects are evidence of an adaptive strategy on the part of the putative adjudicator to promote outcomes that mesh with the adjudicator's own normative (and extralegal) values or preferences, ie those that are supposed to be normatively irrelevant to the legal decision. In addition to the final decisions being influenced by extralegal considerations, we hypothesised that the associated reasons (such as decisions on individual issues, analogies, precedents, and explanations) were also influenced so as to make the final decision appear to be fully coherent and uninfluenced by extralegal factors. Through these rounds of studies, we have revealed a number of novel and, at times, unexpected findings that help to distinguish between the competing theories. These findings seem to be more consistent with the adaptive account than the irrationality explanation.

To summarise, there are a number of findings that can be taken from these

experiments:

First, consistent with previous research, we have replicated the sorts of extralegal effects in the context of a single character manipulation (Studies 1 and 2). We have also elicited these same effects for single character manipulation both at a preliminary indication stage (Study 7) as well as a final decision stage (Studies, 1, 2, 4, 6 and 7).

Secondly, we have shown that these extralegal effects can also be elicited using double character manipulations, where participants are subjected to a character manipulation, followed by a second, opposing, character manipulation before giving a final decision. In these circumstances, extralegal effects are produced in line with the second character manipulation (Studies 6 and 7, though note the lack of statistically significant effects in Study 5 when participants were asked to indicate their leaning at a preliminary stage).

Thirdly, we have shown that these extralegal effects seem to happen only where it is necessary to 'bend the rules' to arrive at a final decision that favours the more sympathetic party (Studies 1, 2, 4, 6, and 7). If the context of the decision is such that the adjudicator can arrive at a decision that favours the sympathetic party by a legitimate application of the rules, they seem to do this instead (Study 3). Thus, where there are both issues that are relevant to character and issues that are not relevant to character, participants seem to appropriately treat character as relevant where it is legitimate to do so, and irrelevant where it is illegitimate to do so.

Fourthly, we revealed that where there was a conflict between appearing to give a coherent decision and favouring the sympathetic party, participants seemed to favour giving a coherent decision (Study 7). Thus, where participants gave a preliminary indication regarding the issues that was impermissibly influenced by a first character manipulation, where a second character manipulation showed the party they favoured to in fact be of opposing character, they would determine their final decision in line with their preliminary indication, even where this effectively punished somebody of good character. The likely reason for the

unwillingness to change their view was that the absence of any relevant information would have made it obvious to observers that they were taking impermissible factors into account. It was also noteworthy that this pattern persisted despite participants being told that their preliminary indication would be confidential to them. Thus they appeared to be sensitive to how they were perceived to observers, ie the experimenters.

Fifthly, we revealed an interesting moderating effect of giving reasons. Whereas giving reasons had little to no moderating effect at the final decision stage (Studies 1, 2, and 6), it had a very strong moderating effect at the preliminary indication stage (Study 6). Thus, though we had seen very clear extralegal effects at the preliminary indication stage if participants did not have to give reasons for their decisions (Study 7), if were required to give reasons at the preliminary indication stage, the extralegal effects disappeared and both groups gave indications that were indistinguishable from each other (Study 6). Thus it seemed that the uncertain status of the evidence, combined with having to give reasons, discouraged participants from impermissibly taking character into account. It is not clear what the mechanism might have been for this, but two possibilities suggest themselves. One possibility is that having to provide reasons made the risk of being caught out by subsequent evidence more salient. The second possibility is that giving the process reasons itself made the participants' decision making more transparent to observers, thereby increasing the risk that they would be subsequently caught out.

### 4.9.2 Limitations

Of course, there are limitations with this research. Given the legal context, it is relevant that participants were lay decision makers and that not all of the decisions were tasks that would be undertaken by a formal court. However, a significant proportion, perhaps the majority of legal adjudication in England and Wales, is undertaken by lay adjudicators in contexts such as the magistrates' courts. Thus Studies 1 and 2 reflected this lay adjudication in both a criminal and civil context and showed extralegal effects. Subsequent studies

focussed on a workplace arbitration and found similar effects. The scenario used in our studies was based on a workplace arbitration used by Simon and his collaborators ((D. Simon, Krawczyk, et al., 2004)), concerning an alleged workplace theft. Just as with Simon's scenario, the factual scenario was considered from the perspective of a workplace arbitration as this slightly less formal context made it more straightforward to introduce the character manipulations. However, the same issues could conceivably be considered by a court or tribunal, for example in civil proceedings for wrongful or unfair dismissal. In a formal legal case, blunt character evidence would often be excluded, but the character of a party is often obvious from more circumstantial evidence such as in Studies 1 and 2. Furthermore, the evidence suggests that many decisions made by lay decision makers in both legal and pseudo-legal contexts generalise to judges and legal adjudicators in legal contexts (Feldman et al., 2016, p. 300; Kelman et al., 1996, p. 303; Lagnado, 2021, p. 121; Leibovitch, 2016; Posner, 2008, p. 248; Schauer, 2010, pp. 103–104; Schauer & Spellman, 2017, p. 261; D. Simon, 1998, pp. 33–34; Spellman, 2010, p. 153). All the same, it would be wise to look to replicate these findings using professional legal adjudicators in distinctly legal adjudicatory contexts.

Related to the above, there was some evidence of what might be seen as inconsistencies in the decisions given by participants in the studies. In particular, while participants often made decisions that favoured a party, they sometimes seemed less inclined to determine the individual issues in that party's favour. Legally, this ought not to have been the case because the decision could not be made in favour of a party unless the issues were also determined in their favour. It seems unlikely that an experienced lay or professional legal adjudicator would have behaved in such a way. It is nonetheless a little bit difficult to classify this as definite inconsistency because of the use of Likert scales for the issues. It would only have been if the issues were determined in a binary way that we could have identified inconsistencies for definite. In any event, given the previous evidence of professional judges taking into account extralegal considerations in decision making, it seems likely that experienced adjudicators would be similarly influenced. This should be tested of course.

A further noteworthy limitation of the research indication is that the manipulations generally relied on an assumed disparity of the general values of a particular legal paradigm and the values of the participants. In many of the paradigms, good or bad character was the information deemed irrelevant in circumstances where the putative party was of such good or bad character that participants appeared to feel compelled to treat this extralegal information as relevant to their decision. However, in doing so, we necessarily assumed that the information was treated similarly by the participants, which may have been the case in the context of the paradigms chosen. Yet, we also know from previous empirical work that there are considerable divergences between adjudicators along dimensions such as political outlook. We also know that there are particular issues such as discrimination that are salient to adjudicators who are personally affected by that issue, and this causes them to make decisions that are different to adjudicators who are not affected by these issues. More sophisticated experimental designs could tease these out. Additionally, for external validity, all experiments were undertaken within the jurisdiction of England and Wales and thereby all the participants recruited were necessarily also resident within the jurisdiction. While such generic issues such as character would be likely to elicit similar responses across different jurisdictions, it is very likely that paradigms raising issues that were more pertinent to participants in particular jurisdictions would likely provoke differing responses.

### 4.9.3 Implications

Of the two contending theories that we have been assessing in the light of the empirical evidence, both imply quite different policy implications. Irrationality type explanations see these sorts of extralegal effects as due to the inadequacy of adjudicators. If adjudicators suffer from such shortcomings, addressing such shortcomings appears theoretically and practically challenging. For those theorists who see these effects as the product of the inherent limitations of human cognition, there seems relatively little that could be done. One implication might be that we should try to avoid reliance on human decision making unless it is strictly necessary. Another implication might be that we provide

adjudicators with some type of cognitive tools to help them avoid making errors in their decision making. An example might be the sorts of causal Bayes nets that can be used to represent thinking in a diagrammatic way such that thinking can be broken down into smaller elements, each of which can be checked individually to iron out logical fallacies (Lagnado, 2021, p. 215). For dual-process theorists, the policy implications are slightly more promising in that they posit that these assumed shortcomings are due to adjudicators using the wrong cognitive system. Thus adjudicators would be capable of avoiding inappropriate decision-making if only they avoided System 1 thinking and used System 2 thinking instead. This could be encouraged either by making adjudicators consciously aware that they were using the 'wrong' system, or by 'nudging' them into using the 'correct' system. This of course is subject to the plausibility of dual process theories, discussed previously at Section 2.8.2 above. If, by contrast, irrationality is not the correct explanation in these circumstances, these effects may be more resistant to the sorts of interventions that irrationality and dual-process theorists recommend.

By contrast, the more rational interpretation that we have been exploring in the course of this thesis points to very different policy implications. We have been examining, as an alternative, whether the extralegal effects that we see are in fact the produce of a strategy that is adaptive or rational for adjudicators in the ordinary legal decision-making environment. That is, in circumstances where an adjudicator wishes to favour a party due to extralegal factors and the decision-making context makes it unlikely that the adjudicator's behaviour will be revealed, a significant proportion of adjudicators will take that factor into account in reaching the decision, and then will present their thinking process as legally justifiable by determining the individual issues and their explanation in accordance with that decision. If this account is correct, some policy prescriptions of irrationality based accounts are less likely to have traction. For example, drawing attention to the problematic behaviour is unlikely to change that behaviour, save insofar as it may make an adjudicator feel there is a risk that their behaviour will be uncovered.

Instead, different strategies are likely to be more fruitful. In the first place it appears

unlikely that extra-legal effects will appear willy-nilly. Instead, the rationality perspective would predict that extra-legal effects will tend to appear where there is a conflict between the formal parameters of the adjudication and the interests of the adjudicator. As other theorists have pointed out, in many areas of law, adjudicators' determinations are fairly consistent with each other (Rachlinski et al., 2017, p. 2051; Sisk & Heise, 2004, p. 746; Sunstein et al., 2006, p. 48). It is only in the more limited handful of areas where there is a diversity of views. The rationality theory would therefore imply that focus would be more profitably be paid to these areas of diversity. Given that the theory posits that extralegal effects occur whenever there is scope for ambiguity about the factors taken into account in making the decision, attenuating this behaviour would require addressing and reducing the ambiguity, primarily by making the adjudicator's behaviour more transparent. Presumably this works because being caught taking extralegal factors into account is a negative blow to the adjudicator's reputation for competence and independence (Engel, 2006; Posner, 2008, p. 81; Thompson, 1985, pp. 429–430). Our research has shown that adjudicators seem very sensitive to the risks of being perceived as having taken extralegal factors into account, even if this means making a decision that has the opposite consequences of what they would otherwise want.

Some particular methods of addressing extralegal effects have been suggested by our prior theoretical analysis at Sections 2 and 3 and seem consistent with previous empirical research. In particular, reason giving seems to have been effective at reducing extralegal effects in many, but not all, circumstances. This would seem to be effective because it makes the adjudicator's otherwise oblique thinking process more transparent. Our research has suggested a further way in which reason giving might be particularly effective in addressing extralegal effects. Participants asked to give reasons at an interim stage exhibited almost no extralegal influences. While the precise mechanism why this is effective requires further clarification, it seems likely that giving reasons either increased the actual risk of participants being shown to have taken extralegal considerations into account or it made the risk of being shown to have done so more salient in the participants' minds. There is an interesting homology with the empirical research in the story model tradition: Pennington and Hastie found that mock jurors assessing the evidence item-by-item came to less extreme decisions

than where the evidence was presented in story order (N. Pennington & Hastie, 1992, p. 201). One explanation for these effects could be that adjudicators or jurors who express an opinion on a piece of evidence thereby 'pin their colours to the mast' meaning that they can only depart from their assessment of that evidence if there is a legal, rather than extralegal, reason to do so. For example, in the context of our studies, a participant who assessed the previous precedent as similar to Tom's case where Tom was of good character could not legitimately depart from this assessment where the only change was that Tom transpired to be of bad character. By contrast, if further evidence was presented that showed that the previous precedent had less in common with Tom's case, a participant could legitimately change their view. It seems that there is some evidence that participants may be aware of this risk which then moderates the extralegal effects otherwise seen.

### 4.9.4 Further Research

In terms of directions for future research, there would be merit in confirming and replicating the effects seen in more naturalistic settings. Thus the experimental context would benefit from being squarely the types of legal decisions made in the law courts. Furthermore, even if much legal adjudication is undertaken by lay decision-makers in England and Wales, it would assist to confirm the robustness of the findings by replicating them with both lay decision-makers with experience in of law courts and tribunals as well as with professional adjudicators.

Given the quite generic nature of the extralegal information manipulated in these experiments that assumed a fairly homogenous response by participants, it would provide further support for the theory outlined here if more sophisticated experiments were designed to elicit responses that depended on the individual values, both personal and cultural, of the participants. Within jurisdictions, matters such as the political outlook of the participant could be taken into account when designing paradigms that highlighted the salience of political outlook, and across jurisdictions, paradigms could be designed that raised issues that were

more or less salient depending on the culture of the participants.

The findings from our research together with previous research suggest that extralegal effects are fingerprints of quite sophisticated strategies by adjudicators that are also sensitive to quite a range of factors. Teasing out the full contours of the empirical picture will require further empirical and theoretical work. Given that these sorts of extralegal effects are by definition unacceptable in legal decision making, one important consideration would appear to be the existence of an observer who could set in train the sorts of consequences that would be unwelcome to the adjudicator. Examples might include overturning the decision or adverse consequences for the adjudicator themselves such as a loss of status with either negative consequences for them individually or for their ability to influence future adjudications. As we have seen, the appearance of consistency appears to be valued highly by participants, even higher than the outcomes in an individual case. There is some sense in this in that loss of reputation could risk the ability to exercise influence in multiple future cases. Thus sacrificing influence in one case to preserve influence in multiple cases appears logical.

The giving of reasons also appears to be the key method of making the decision process of adjudicators more transparent to observers, and thereby addressing extralegal effects. But as our research has shown, there is more to the story than simply giving reasons or not. We saw that there appears to be a relationship between giving reasons and the point at which the reasons are given. For participants, giving reasons at an inchoate stage when what future evidence might be presented seems to have more of a moderating effect on extralegal effects than were reasons are expected once the evidence is all complete. We have speculated on the reasons for this, but further research and theory is needed to clarify exactly why this happens.

Finally, while uncertainty seems a necessary ingredient for these types of extralegal effects to occur, greater formality about the nature of this uncertainty would be helpful to better understand the phenomenon. At one level there seem to be matters that make it quite obvious in an individual case that that an adjudicator has behaved improperly: an example is

where there are clear counterfactual cases that differ only by the factor that adjudicators are not supposed to take into account. If adjudicators come to different decisions in the two cases, it is a reasonable inference that they have taken into account that extralegal factor. However, between the certainty of this extreme and the opposing extreme where it is difficult to infer anything about what the adjudicator took into account there are various shades of grey. Future research could constructively illuminate the level and type of uncertainty that encourages and discourages these extralegal effects.

# 5. THE EFFECT OF CASE PRESENTATION ORDER ON OUTCOMES

## 5.1 INTRODUCTION

In Study 7 of the previous section, we saw that participants arrived at very different decisions depending on whether or not they had previously given a preliminary indication on the same issues. This suggests a potential link to the parallel research domain of moral decision making, where it is well established that participants often make different decisions depending on whether or not they have previously made a decision on related issues. A common paradigm is for participants to resolve pairs of similar moral dilemmas, often where one dilemma is generally approved of and the other is disapproved of, in a randomised order. The pattern commonly seen is that participants' decisions in the dilemma seen second are more likely to reflect the decision taken in the dilemma seen first than if the second dilemma had been viewed in isolation (Schwitzgebel & Cushman, 2012, pp. 141–142). For example, many of the dilemmas are based on a theme of runaway trams or (adopting the US terminology) 'trolleys' that will hit and kill different people depending on how the decision maker intervenes. Often there are five people who will die if the trolley continues to travel along its trajectory, but only one person who will die if the participant diverts the trolley to another track (Schwitzgebel & Cushman, 2012, p. 138). In 'Switch' a scenario where the participant can divert the trolley by pulling a switch, most participants approve of intervention (Appiah, 2008, p. 89; Nichols & Mallon, 2006, p. 531; Unger, 1996, p. 87). But by contrast, in 'Fat Man', a scenario where participants can prevent the killing of the five people by pushing a fat man off a bridge into the path of the trolley (to his death), most people disapprove (Appiah, 2008, p. 89; Nichols & Mallon, 2006, p. 531). Most pertinent for our purposes is that Switch tends to be approved if assessed first or in isolation, but disapproved of if assessed after Fat Man (Lanteri et al., 2008, p. 796; Lombrozo, 2009, pp. 281–282; Norcross, 2008, p. 67; Schwitzgebel & Cushman, 2015, p. 131; Wiegmann et al., 2012, p. 816). While these types of decisions are considered moral decisions, it is not difficult to see that decisions that involve life or death consequences could readily amount to tortious or criminal decisions in a legal context. Yet while this research programme has triggered

enormous empirical and theoretical work in the field of morality (Appiah, 2008, pp. 90–91), there historically been a relative dearth of research in the related legal field (Lindquist & Cross, 2005, p. 1173; Spamann et al., 2021, p. 113).

This paucity of research is somewhat surprising given the apparent importance of consistency and precedent in the legal context. It is noteworthy that the fields of morality and law seem to take very different attitudes to these sort of order effects. In the field of morality, order effects are seen as deeply problematic. Moral philosophers often assume that moral decisions ought to derive from the application of a stable set of values or principles to the facts(Schwitzgebel & Cushman, 2012, p. 136) and that therefore any variation when faced with the same set of facts is a sign of irrationality (A. B. Moore et al., 2008, p. 556; Rini, 2015, p. 438; Sinnott-Armstrong, 2007, pp. 54, 67). Despite this view, these order effects seem to affect both lay decision makers just as much as professional moral philosophers, and they also seem resistant to common debasing techniques (Schwitzgebel & Cushman, 2012, p. 147, 2015, p. 128). By contrast, a more accommodating approach seems to be taken in the adjudicatory context where there is a recognition that order effects may be appropriate in some circumstances. This approach stems from a variety of factors. For one, there is a general recognition that finding an appropriate decision to resolve a contested family of legal disputes may be challenging and there may be a number of solutions that are reasonable. For example, with the advent of reliable postal services an issue arose as to whether a contract acceptance in a letter was effective when it was sent or when it was received. Initially, some jurisdictions preferred the former solution whereas other jurisdictions preferred the latter. Similarly, different jurisdictions have taken different approaches to the doctrine of necessity as a defence to murder. In England and Wales a rule has developed whereby necessity is never available as a defence to murder, whereas in the United States it is available. Relatedly, there is a recognition that legal decision making depends to some extent on the particular adjudicator. Secondly, there is a recognition that law has practical consequences: certainty and predictability of a court may matter more to potential litigants than perfection. Thus an imperfect rule that is applied consistently may be preferred by litigants to inconsistent application of different rules while the court aspires to a perfect rule (J. H. Baker, 2002, p.

199; Engel, 2006, p. 225). Others see copying previous precedents as a way of making a court's task less onerous (Heiner, 1986, p. 236). In common law jurisdictions this approach is recognised by the Latin phrase *stare decisis* or 'letting the decision stand' (M. Shapiro, 1972) and a similar approach is taken in civil law jurisdictions (Posner, 2008, p. 145). Consequently, order effects whereby courts in subsequent cases adopt the approach taken in earlier cases are acceptable and often commended in the legal context.

Order affects are foreseen in law in various guises. One of the most noteworthy is Dworkin's 'Chain Novel' theory whereby the common law is seen as a novel written *seriatim* by a series of authors: each court is seen as like an author who has some leeway to develop the story as they see fit, but is bound by the characters and situations established earlier in the book by other authors (Dworkin, 1986, pp. 228–238). Other theorists, particularly those in the law and economics tradition, speak of 'path dependence' (Kornhauser, 1992). In other words, it is assumed that the law develops along different paths depending on the order in which the courts resolve different cases (Lindquist & Cross, 2005, p. 1169).

### 5.1.1 Empirical Research

Given that order effects are envisaged in legal decision making, but not in moral decision making, it is somewhat ironic that most of the empirical evidence for order effects is in the latter field. Much is, as noted previously, in the context of moral 'trolley' dilemmas (Lanteri et al., 2008, p. 796; Liao et al., 2012, p. 664; Lombrozo, 2009, pp. 281–282; Nichols & Mallon, 2006, p. 536; Norcross, 2008, p. 67; Nucci, 2013, p. 667; Petrinovich & O'Neill, 1996, p. 156; Wiegmann et al., 2012, p. 816). Thus responses to Fat Man remain similarly unacceptable regardless of the cases that precede it (Wiegmann & Waldmann, 2014, p. 28), but responses to the Switch dilemma vary considerably depending on whether Switch is presented before or after certain other dilemmas. This effect is observed both with homogeneous and heterogeneous paired dilemmas (Sinnott-Armstrong, 2007, p. 62). Thus Switch also varies considerably depending on whether or not it is presented after 'Transplant',

a heterogeneous scenario where participants are asked whether they would kill one healthy person in order to transplant their organs to save the lives of five others (Norcross, 2008, p. 67). Unsurprisingly almost all participants object to the Transplant scenario, confirming the finding that some scenarios such as Switch are more labile than others such as Transplant (Petrinovich & O'Neill, 1996, p. 155; Schwitzgebel & Cushman, 2015, p. 131; Wiegmann et al., 2012, p. 816). As a result, some order effects are fairly asymmetric, with one scenario being labile with the other stabile (Schwitzgebel & Cushman, 2015, p. 128; Wiegmann & Waldmann, 2014, p. 29).

In the legal field there has been some research into order effects, but these have generally been into the effect of different evidence presentation orders within a single case. Thus researchers have examined the effect of presenting evidence that favours a particular party at different points within the same case, such as at the start or at the end of the case. The evidence has been somewhat mixed, with many finding primacy effects, but some finding evidence of recency effects. Anderson (1959) found a recency effect when evidence was presented to participants in the context of a mock jury trial, as did Wilson (1971). Subsequent replications by Furnham (1986, p. 355) were consistent with this research. Likewise, Costabile, & Klein (2005, p. 47) found a similar pattern, putting it down to the effect of mock jurors being able to better remember the incriminating evidence when it was presented last. More recent research has also found some support for a recency effect (Engel et al., 2020). However, studies with more realistic legal materials, on undergraduate students, have found evidence for a primacy effect (D. C. Pennington, 1982) and other research has found both primacy and recency effects depending on the manipulations used (Kerstholt & Jackson, 1998, p. 445). Nonetheless, findings from the order in which evidence is presented within a case are not quite akin to the effect of different presentation orders of cases on decisions.

More closely related to the question we are interested in is the phenomenon of anchoring in legal cases, particularly where the 'anchor' consists of a previous legal precedent (Bordalo et al., 2015, p. S25). Anchoring is where outcomes are influenced by a seemingly irrelevant value and the phenomenon has been widely demonstrated within decision-making

generally (Bahník et al., 2022; Chapman & Johnson, 1999; Epley, 2004; Epley & Gilovich, 2001; Mussweiler & Strack, 1999; Northcraft & Neale, 1987; Tversky & Kahneman, 1974). Thus a participant asked to estimate a value will often be influenced to choose a higher value when they have been shown an arbitrarily generated value immediately prior to giving the estimate and *vice versa*. These effects seem to influence experts as much as non-experts (Englich et al., 2006, pp. 193–194). Anchor effects also seem particularly likely where the value to be estimated is uncertain (Feldman et al., 2016, pp. 306–307). This suggests a possible link with the order effects seen in moral trolley dilemmas where dilemmas generating a greater diversity of responses such as Switch tend to be more likely to be influenced by preceding cases than those where responses are almost unequivocal such as Transplant. Given the widespread evidence for anchoring generally, it is unsurprising that it has been shown to occur in the legal domain too. Thus where an adjudicator has to select a value such as an amount of damages or a length of sentence, arbitrary anchors have been shown to be influential. For example, the amount requested by the claimant in a personal injury case has been consistently shown to be significantly correlated with final awards (Feldman et al., 2016; Guthrie et al., 2001; Hastie et al., 1999; Malouff & Schutte, 1989; Marti & Wissler, 2000). Interestingly, some have also found that the anchor may affect both the final amount of damages awarded and the likelihood of a defendant being found responsible for the injury (Chapman & Bornstein, 1996). Similarly, Enlich, & Mussweiler (2001) found that legal professionals playing the role of judges were strongly influenced by the demands of the prosecutor, a pattern that seems to be replicated in the real world (Dhami, 2003; Englich et al., 2005). Just as in psychology generally, anchors seem to be effective even when obviously randomly generated (2006, pp. 192–193). While most of these anchors were in the form of arbitrary values presented before or during a case, Feldman et al conducted an experiment based on a copyright infringement based on unauthorised sampling of 8% of another's music (Feldman et al., 2016). Anchors were provided in the form of previous precedents where a 1% sample was held never to to amount to a breach and another where a 50% sample was held to always amount to a breach. These precedent anchors had a significant influence on whether participants found the sample in the target case to be a breach of copyright.

The most pertinent research for order effects is that which has actually examined the effect of previous precedents, particularly precedents with factors beyond simple low or high anchors. In doing so, it is helpful to distinguish between vertical and horizontal precedent. In most jurisdictions there is a hierarchy of courts and more junior courts are expected to follow the decisions of senior courts, a relationship that can be described as vertical precedent. In common-law jurisdictions, principles established by senior courts have a similar status to legislation. Thus, a junior court would generally be expected to follow a decision of a senior court. By contrast, there is less of an expectation that courts should always follow the previous decisions of a court at the same level, a relationship that can be described as horizontal precedent. It is in the context of horizontal precedent that order effects would be most interesting. Nonetheless, there has been relatively little research in this area (Lindquist & Cross, 2005, p. 1173; Spamann et al., 2021, p. 113). One exception is research looking at the effect of judicial attitudes in cases of first impression. Cases of first impression are those where the adjudicator identifies that there are no relevant precedents, the implication being that there is less restriction on the possible decisions that the adjudicator might make. Consistent with the presumed influence of precedent, Lindquist & Cross (2005, p. 1156) found that in real-life US cases, a judge's attitude was more influential in cases of first impression. Similarly, Spamann et al (2021) gave almost 300 professional judges from different jurisdictions an international criminal law case to determine, and found a small effect of horizontal precedent on decisions on guilt. Likewise, Simon found evidence that participants could be influenced by character to determine an analogy in a case one way or the other and, once they had done so, they determined the analogy in the same way when asked to decide the same analogy in the context of a second case dealing with a very different subject matter (Holyoak & Simon, 1999, pp. 11–18).

### 5.1.2 Theoretical Explanations

Some popular theories to explain these types of order effects, both in the moral and legal domains, are based on the idea of consistency. Within consistency explanations, there are two further subcategories of explanation that can be termed institutional and individual (Lim, 2000). Institutional consistency theories are more associated with the legal domain, for they assume that the reasons for the order effects are to promote wider societal benefits such as certainty and predictability for litigants. This is in line with accepted rationales for the principle of *stare decisis*. Thus, a court that would prefer a different outcome if it was in the context of a case of first impression might nonetheless take a different decision if it accords with a previous precedent. This affords litigants some degree of predictability consistent with the idea of the rule of law (Lindquist & Cross, 2005, pp. 1159–1160). By contrast, individual consistency theories are associated with both the legal and moral domains. The basic idea of individual consistency is that there is an advantage to the individual adjudicator in being consistent (or, correspondingly, a disadvantage in being inconsistent) (Lim, 2000, p. 723). Thus Unger says 'folks want their responses to seem consistent' (Unger, 1996, p. 92) and Schwitzgebel, & Cushman refer to a 'general desire to maintain consistency in judgment' (Schwitzgebel & Cushman, 2012, p. 148). Some theorists point to the similarities between scenarios as a relevant factor in promoting consistency (Horne & Livengood, 2017; Petrinovich & O'Neill, 1996, p. 156; Unger, 1996, pp. 93–94; Wiegmann et al., 2012, p. 819), though others observe that order effects are also seen across heterogeneous scenarios (Sinnott-Armstrong, 2007, p. 62). Nonetheless, such speculations do not always explain why individual consistency or inconsistency matters. One possibility links to that which we explored in Section 3 is that individual inconsistency may be taken by observers as evidence of partiality or incompetence (Engel, 2006, p. 250; Posner, 2008, p. 81; Thompson, 1985; Walton, 2005, p. 48). This would give adjudicators a motive to avoid this impression. Some research has compared institutional against individual consistency. Brenner and Spaeth examined dissenting views on the US Supreme Court and found that dissenters in appeals generally maintained their dissent in subsequent cases when the issue came before the court again, rather than following the majority in the earlier case which by then had become

precedent (Brenner & Spaeth, 1995, p. 287). This would suggest a desire to be consistent outweighed a desire to follow precedent.

One issue with consistency based explanations is in explaining asymmetrical order effects whereby one case in a pair affects the other but not vice versa. One idea is that some dilemmas such as Transplant provoke a high degree of consensus, making departure from this consensus as less attractive to participants than attempting to give an impression of consistency. Again, there are links with the idea of the 'zone of reasonableness' available to decision makers discussed in Section 2. However, some theorists take a different tack to explain the asymmetry. Thus, Ortony argues that Scenario A can be more like Scenario B than Scenario B is like Scenario A due to aspects of A that are more salient than B (Ortony, 1993). However, as Wiegmann & Waldmann point out, reliance on the concept of salience seems somewhat unconstrained (Wiegmann & Waldmann, 2014, p. 40), akin to reliance on a Molieresque *virtus dormitiva*. Wiegmann & Waldmann offer a more nuanced explanation based on the ambiguity of the underlying causal structure, a consideration that may be particularly germane to the structure of trolley problems. The idea is that the ambiguity of some scenarios, and their associated susceptibility to influence is because their underlying causal structure is less clear-cut (Wiegmann & Waldmann, 2014). Thus ambiguous scenarios are affected by unambiguous scenarios but not vice versa.

A slightly different type of explanation suggests that scenarios that influence other scenarios do this because they highlight factors that are salient to the decision that the decision-maker may not have taken account of until they were explicitly or implicitly highlighted (Schwitzgebel & Cushman, 2012, p. 149). Legal theorists might find such explanations appealing given that there is a recognition that identifying a principle to account for all the different cases that are likely to come before the courts can be challenging, in part because it can be difficult to predict what cases will come forward (Hart, 1961, p. 128). Schwitzgebel & Cushman themselves reject such an explanation, in part because the order in which cases are presented seems to affect the subsequent principles that a participant is willing to endorse even after they have seen all the cases (Schwitzgebel & Cushman, 2012, p.

149).

### 5.1.3 Our Research

In the following experiments, we sought to replicate the type of order effects commonly seen in moral decision-making dilemmas in a legal context using both criminal and civil law examples. Translating the experimental design from the moral sphere to the legal sphere was facilitated due to the binary nature of legal cases, which often have two parties and binary outcomes (such as guilty or not guilty, or to grant or refuse the application). Similarly, we initially adopted the approach taken in moral trolley cases of presenting two apparently homologous cases, but where one case was normally approved of in isolation and the other case was disapproved of in isolation. Participants were presented with the two cases in a randomised order. We hypothesised that we would see the same types of order effects as seen in moral decision making where participants responses in the dilemma seen second would move closer to responses in the dilemma seen first, and that it was likely that such effects would be asymmetric with one dilemma being much more labile than the other in the pair.

## 5.2 STUDY 8

In Study 8, the underlying legal principle to be determined by participants was whether it was lawful for an individual to commit suicide on hospital premises so as to facilitate that individual becoming an organ donor. In accordance with the law in England and Wales, it is no longer an offence for an individual to commit suicide, but it is an offence for an individual to assist an individual to attempt or to commit suicide (ss.1 and 2 Suicide Act, 1962). In situations where the lawfulness or otherwise of a course of action is in question, the medical authorities may seek a declaration from a civil court in advance to ask for guidance.

For example, at a time when it was legally uncertain whether the removal of treatment from a minimally conscious patient, the hospital trust treating him sought guidance from the High Court of England and Wales as to whether the discontinuance of medical treatment (save for purposes such as reducing pain) would be lawful (*Airedale NHS Trust v Bland*, 1993). The desire for those intending to commit suicide to also become donors has been reported in the media and academic literature (Allard & Fortin, 2017; Anonymous, 2013; Miller, 2017), the concept of using organs from individuals committing suicide has been discussed in the academic literature (Bollen et al., 2017; Dijk et al., 2018; Shaw, 2014; Wilkinson & Savulescu, 2012), and such donation takes place in jurisdictions such as Belgium and the Netherlands where assisted suicide is legal (Bollen et al., 2016; Gilbo et al., 2019; Van Raemdonck, 2011; Van Raemdonck et al., 2011; Ysebaert et al., 2009). As such, it was considered a reasonably plausible, if unlikely, legal scenario, correspondingly with limited or no precedent.

Five different scenarios involving taking difficult legal decisions in order to save a greater overall number of lives were pretested to identify two similar scenarios that participants found differently acceptable. Of these, two scenarios based on the above case were selected for the main experiment. In the first, the reason that the individual wished to commit suicide was because they were suffering from multiple sclerosis ('MS'), a long-term incurable condition which the symptoms can be addressed. When presented with this variant, most pre-test participants agreed that the proposed course of action was acceptable (82% approval v 18% disapproval, n=11). In the second scenario, the reason the individual wished to commit suicide was because they suffered from long-term depression and most participants in the pre-test instead objected to this (42% approval v 58% disapproval, n = 12). Of the five scenarios pretested, the long-term depression scenario was the only one that had an overall disapproval rating.

Our main hypothesis was that we would see order effects such that participants' responses in the scenario seen second would be closer to their response in the first scenario compared to where that second scenario was seen in isolation. For example, while responses

in the MS scenario were generally acceptable when the scenario was presented in isolation, we predicted that responses to the MS scenario would be seen as less acceptable when presented after the generally unacceptable depression scenario, and the opposite would be seen for the depression scenario. Given widespread previous evidence of an asymmetrical effect, we envisaged that it was possible that responses in one scenario would be more labile than in the other, but we had no theoretical reason to predict which would be the most labile.

## 5.2.1 Method

### 5.2.1.1 Participants

One hundred and ninety-nine participants recruited using the online survey platform Prolific completed the survey (124 females, 75 males; aged 40 to 65, M age 50.3, SD 7.4; 2.5% were students; 45% in full time employment, 26% in part time employment, 29% unemployed or other) and were paid £0.50 for their time. The design was calculated to have an 80% power to detect an effect size of d=0.40. Participants were selected on the basis of British nationality and residency in England and Wales. The study was conducted in accordance with approval obtained from UCL's Research Ethics Committee (EP/2018/005). Informed consent was obtained from each individual in advance of participation by providing them with study information and a consent form to agree to.

### 5.2.1.2 Design

We used a 2 × 2 mixed design in which all participants viewed both cases, the depression scenario and the multiple sclerosis scenario, sequentially but in a randomised order. Given that all participants viewed both scenarios, this was a within subject manipulation, whereas participants only viewed the scenarios in one of two possible orders,

hence order was a between subject manipulation.

#### 5.2.1.3 Materials

The materials were hosted on the online platform Gorilla Experiment Builder (www.gorilla.sc). The case materials consisted of two summaries of the facts of the case and the issue to be decided. Both summaries were essentially identical other than the underlying cause of the individual's condition. In each case, participants were told that there were two people in need of urgent organ transplants who would die if they did not receive a transplant soon. However, an individual, hospitalised with the relevant condition, would be a suitable donor. Participants were told that, despite undergoing treatment, this individual wanted to die and also wanted to donate his organs. Participants were told that because of this, the individual wished to commit suicide in a hospital so that the hospital could receive his organs in a good condition.

In each case, participants were advised that while it was legal to commit suicide, it was not legal to assist an individual to commit suicide, and, for this reason, the individual and the hospital were making a joint application to the court to seek guidance as to whether it was lawful for the individual to commit suicide on the hospital premises.

Once they had read the information in each scenario, participants were asked to imagine that they were the judge and to make a decision on the acceptability of the particular application. Participants were first asked whether they would or would not grant the application, then asked secondly to indicate on a percentage scale from 0% to 100% the acceptability of the application, and then thirdly to explain the reasoning behind their decision.

#### 5.2.1.4 Measures

As noted above, participants were asked to indicate three measures. For reasons of external validity, participants were asked to give a ruling on whether they would grant or oppose the application, and their reasons for this. These would be matters that would ordinarily be expected of a real-world adjudicator, save that the reasoning of a High Court judge would be likely to be much more detailed than that expected of our participants. In addition, both to provide more insight into participants' cognition, as well as to be consistent with the measures typically taken in trolly type moral decision making research, participants were also asked to indicate their view of the acceptability of the application on a 100 point percentage scale where the extremes of the scale represented completely unacceptable and completely acceptable.

#### 5.2.1.5 Procedure

As previously noted, participants were recruited online and participated in the survey using the online platform Gorilla Experiment Builder in a place of their choosing, using their own device. On referral from the Prolific platform, participants were first provided with the study information form. They were then asked to complete the consent form comprising a number of statements to which an affirmative answer was required in order to participate in the survey. An anonymous user identification was collected to enable subsequent matching of demographic data without compromising the participants' anonymity.

Participants were randomly assigned to one of the two conditions by the Gorilla Experiment Builder such that half of participants viewed the MS scenario followed by the depression scenario and half viewed the depression scenario first followed by the depression scenario. Participants were required to give responses to the first scenario viewed before they could complete the second scenario. After reviewing the first scenario to which they were assigned, participants were asked to indicate whether they would grant or oppose the application, to indicate how acceptable they found the application on a 100 point % scale

from 0% to 100%, and were asked to explain the reasons behind their decision. Once they had completed the first scenario, they were then presented with the second scenario and asked to complete the same measures.

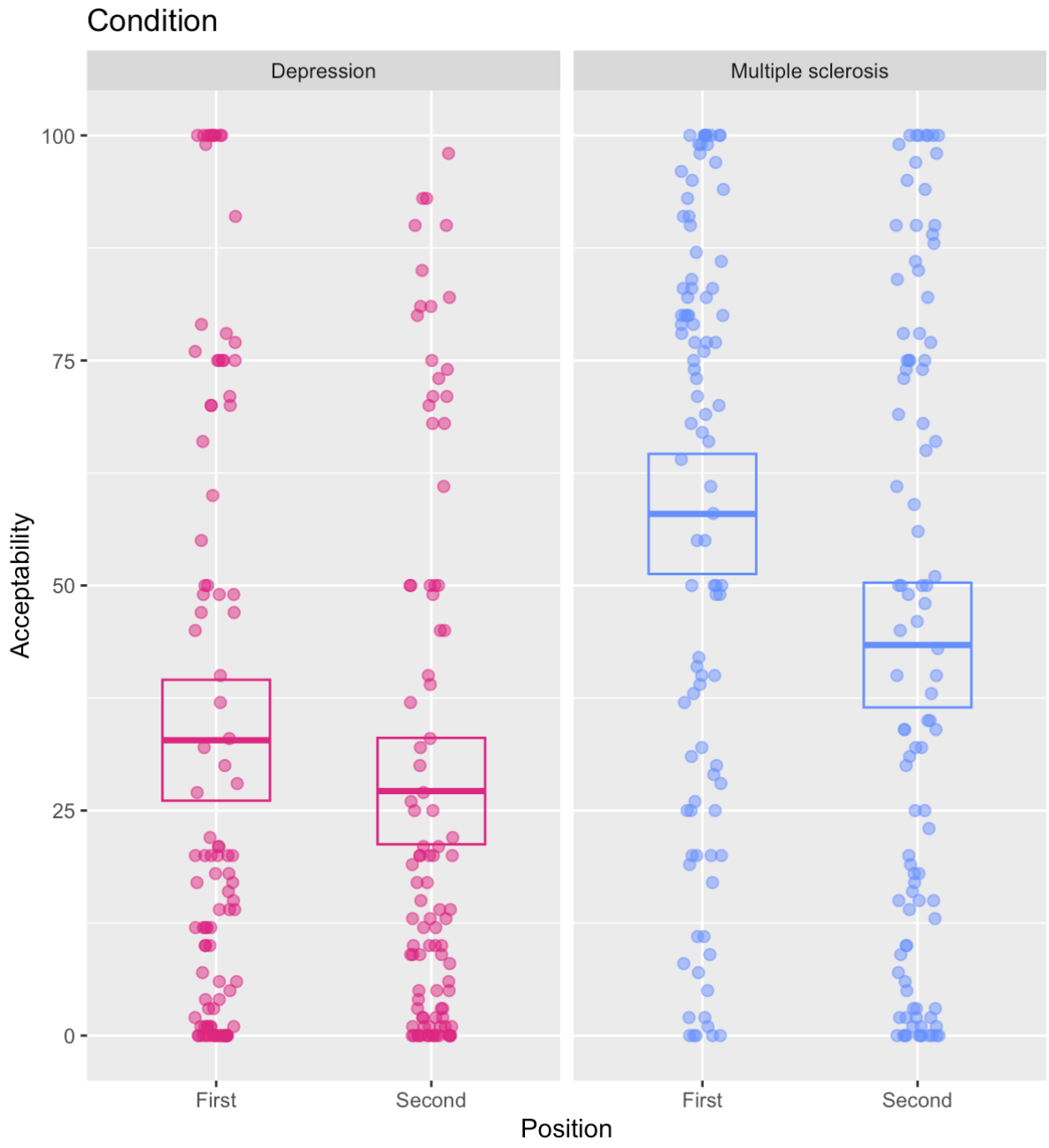
After completing the survey, participants were thanked for their participation and referred back to the Prolific survey platform to confirm their participation. Once both platforms had confirmed their participation, their remuneration was authorised.

### 5.2.2 Results

When considered first, and therefore in isolation, participants found the depression application less acceptable (M=32.8%, SD=34.0, n=101) than the MS application (M=57.9%, SD=33.3, n=98), consistent with pre-testing. A two-sample t-test confirmed that this difference was statistically significant at the 0.05 level ($t(197)=5.27$, $p<0.001$, 95% CI [15.73, 34.55], Cohen's D=0.75). Correspondingly, participants were less minded to grant the depression application (80:21 refuse:grant) than the MS application (37:61 refuse:grant). A Fisher's exact text similarly confirmed that this difference was statistically significant ($p<0.001$, OR 6.21, 95% CI [3.20, 12.45]).

As predicted, there was an order effect and as foreseen, the depression scenario was not labile whereas the MS scenario was, see Figure 11. Thus, the depression scenario remained generally unacceptable whether seen in either in isolation or after the MS scenario, even becoming slightly more unacceptable when seen in second position (32.81% v 27.16%). However, this small difference did not reach statistical significance ($t(197)=1.25$, $p=0.21$, 95% CI [-3.27, 14.56], Cohen's D=0.18). Correspondingly, participants generally refused the application whether it was seen first (80:21 refuse:grant) and were slightly less likely to grant it when it was reviewed after the MS scenario (89:9 refuse:grant). A Fisher's exact text indicated that this difference was significant ($p=0.029$, OR 0.39, 95% CI [0.15, 0.95]).

*Figure 11. Mean acceptability ratings by condition and scenario position from Study 8.*




By contrast, responses to the MS scenario changed much more drastically, flipping from general approval to general disapproval, as shown in Figure 11. The MS scenario

received a mean acceptability well over 50% when seen in isolation (M=57.9%, SD=33.3, n=98), but this dropped to well below 50% when reviewed after the depression scenario (M=43.4%, SD=35.1, n=101). A two-sample t-test confirmed that this difference was statistically significant (t(197)=3.00, p=0.003, 95% CI [5.00, 24.13], Cohen's D=0.43). Unsurprisingly, decisions to approve or reject the application changed similarly, such that most participants granted the MS application when seen in isolation (37:61 refuse:grant) but refused it when it was reviewed after the depression scenario (62:39 refuse: grant). A Fisher's exact text was consistent with this (p=0.001, OR=0.38, 95% CI [0.21, 0.70]).

### 5.2.3 Discussion

The headline finding of this study is that it replicated the types of order effects seen in the moral decision making domain in a new domain, that of legal decision making. Thus, when viewed in isolation, one of the legal dilemmas presented to participants, the MS scenario, was found to be both generally acceptable to participants on the continuous variable that measured acceptability and to be generally favoured on the categorical variable that measured whether participants would grant or refuse the application. However, when the MS scenario was presented after another similar scenario, the depression scenario (that most participants contrastingly found unacceptable and refused the application), the picture was very different. In this context, participants found the scenario to be generally unacceptable and generally disfavoured the application by tending to reject it.

In addition, the results also indicated a pattern that is very characteristic of trolley type experiments in the moral decision making research field of asymmetric order effects with one of the scenarios being significantly more labile than the other. Here, it was the depression scenario that was stabile. When seen in isolation, the depression scenario was both generally disapproved of and correspondingly the application was generally refused. When seen following the MS scenario (which was generally approved of, at least when viewed in isolation), responses to the depression scenario hardly changed, with participants again

generally finding it unacceptable and refusing the application. While there was some evidence of a difference in responses when the scenario was seen after the MS scenario, this was in the opposite direction to that predicted by all theories (ie, the depression scenario was assessed as slightly less acceptable when presented after the generally acceptable MS scenario) and the difference was only statistically significant on one of the two measures, the categorical variable of whether the application should be granted or not.

These empirical findings seem to chime with consistency or coherence based explanations in that those theories assume that legal decision makers, when faced with similar cases, feel some impetus to treat those cases alike. Individual consistency theories would assume that the decision maker feels this impetus personally as failure to do so might be perceived negatively by observers in that it could be taken as evidence of carelessness, partiality, or other improper motive. Institutional consistency theories would assume that the decision maker feels this impetus due to expectations on the court as an institution to follow previous precedent in order to maintain predictability for litigants and members of society. Given that our paradigm only provided a previous precedent from the same participant, it is difficult to distinguish further between individual or institutional type theories.

At the same time, the asymmetry of the empirical findings, with one scenario apparently being influenced but not the other, does not seem wholly compatible with consistency type explanations, even if this pattern is often seen in the moral decision-making context. To the extent that there is some impetus for a decision-maker to be consistent, one might assume that it would be in both directions: given that a participant cannot change the assessment and decision that they arrived at in the first scenario, the only way to appear more consistent would be for their response in the second scenario to be more akin to the first. This would imply that responses in the depression scenario should become more acceptable when it follows the more positively perceived MS scenario; but as we have seen responses remain stable, even slightly more negative. One possible explanation for this recognises other influences on decision-making and links back the idea of a decision maker's 'zone of reasonableness' as discussed in Sections 2, 3, and 4. As previously discussed, theorists such

as Posner have recognised that collateral or extralegal influences on decision making can be moderated by other factors. A decision-maker can take these extralegal influences into account provided there is some leeway in the range of possible decisions they can take that will appear reasonable to an observer. In some contexts this zone of reasonableness will be wide, and in others it will be narrow. For example, in the 'transplant' moral decision-making scenario discussed above, anything other than condemnation of the proposed murder and distribution of the victims' organs will appear unreasonable and thus a decision-maker's zone of reasonableness will be extremely narrow. In the context of our depression scenario, one explanation for its stability is that any impetus to be consistent is constrained by a narrow zone of reasonableness.

In terms of other explanations, salience theories suggest that a scenario may influence another through highlighting important aspects of the decision that the decision-maker may have otherwise not considered. For instance, in our dilemmas, the depression scenario may have highlighted issues such as the possibility of treating or alleviating the symptoms of the condition, or 'slippery slope' type arguments that permitting those committing suicide to donate organs in some situations might lead to people being permitted or implicitly encouraged to take this course of action in more trivial circumstances. Finally, theories focusing on the relative ambiguity of the underlying causal structure of the scenarios seem less pertinent here, given the relative unimportance of causal structures in our dilemmas compared to those in familiar trolley-type problems.

While these findings are interesting and important in that they demonstrate order effects from sequential decision-making in the new domain of law, some caution needs to be exercised against drawing too conclusive inferences from such limited evidence. Further replication in the legal context is probably the first priority. At the same time it is important to recognise the limitations of making generalisations from inexperienced lay decision-makers to experienced or professional decision makers. The limited paradigm used for this experiment is not nearly as formal or detailed as the real-life context. In an equivalent real-life case, there is likely to be considerably more documentary and witness evidence that may

be tested in cross-examination. The decision-maker will also be assisted by counsel for the parties and possibly by an *amicus curiae* or friend of the court offering wider arguments. Though participants here were asked to give reasons for their decisions, these reasons would not be comparable to a formal legal judgment from a court. Still, it makes sense to establish empirical patterns using lay decision-makers before seeking to replicate those findings with the more high-stakes context of professional judges, and findings established using lay decision-makers invariably generalise to the professional context.

Another dimension to this study with implications for further study is the limited extent to which it distinguishes between different theoretical explanations for these phenomena. As noted above, there is no strong basis to prefer coherence or consistency type theories (whether individual or institutional) over salience type theories, though causal type theories do not seem particularly relevant given the particular paradigm used. As such, a secondary consideration for future research should be to start seeking to distinguish empirically between these types of theories.

## 5.3 STUDY 9

Our second study of this section sought to replicate the findings of Study 8, but in the different context of criminal rather than civil proceedings. We again sought to use the standard paradigm of presenting two cases sequentially where one case was generally assessed favourably in isolation while the other case was generally assessed negatively in isolation. However, because penal decisions are primarily about the guilt or innocence of an accused rather than decisions that had consequences for lives lost or saved, the context was necessarily slightly different. As a result, we asked participants to rule whether an accused was guilty or not guilty after making a decision that had lead to loss of life. The inspiration for the scenarios was a relatively old but notorious common-law case of *R v Dudley v Stephens* (1884) 14 QBD 273 DC. In the real-life case, two sailors who had been shipwrecked survived by killing and eating a cabin boy. They were subsequently prosecuted

for murder. Somewhat unusually, the trial judge decided to ask the jury to return a 'special verdict' confined to the facts, reserving the question of law as to whether these facts amounted to a crime to be determined by judges. In the case the judges subsequently ruled that the defence of necessity, also known as duress of circumstances, was not available to murder, and found both accused guilty of murder.

As with our Study 8, we selected two versions of this scenario with contrasting levels of acceptability. The less acceptable version ('apprentice') was very similar to the facts in the original case of *R v Dudley v Stephens*: three sailors were shipwrecked and surviving on a life-raft. Two of the senior members who were starving, the captain and the first mate, unilaterally attacked and killed a more junior member, and consumed him. In pre-testing, this scenario was found to be generally objectionable. The mean acceptability was relatively low at 39.4% on a 100% scale from completely unacceptable to completely acceptable and, correspondingly, verdicts were generally guilty (guilty:not guilty = 6:2). By contrast, in the second - more acceptable - version ('captain') the three crew members discussed how to proceed and agreed to draw lots to decide who would be sacrificed. The captain happened to be the the one who drew the short straw, and he allowed the mate to kill him before the two survivors consumed him. In pre-testing, this was seen as more acceptable (76.2%) and correspondingly verdicts tended to be not guilty (guilty:not guilty = 3:9). As with the first study, the facts were not in dispute, so the participants were not required to engage in fact-finding, only decision-making. As before, each participant considered both scenarios in a sequential order with the order or presentation being randomised between participants.

Given previous research in the moral decision-making contest and our findings from Study 8, our primary prediction was for an order effect such that responses in a scenario seen second would tend to be closer to responses to the scenario that preceded it than if that scenario was seen first (ie, in isolation). Given previous findings, we also foresaw as a possibility that responses would be asymmetrical, such that responses in one scenario would be more labile, with responses in the other scenario would be more stabile. Again, given that these scenarios were novel, we had not particular theoretical basis to select which scenario

would be labile and which would be stabile.

## 5.3.1 Method

### 5.3.1.1 Participants

Two hundred and ten participants recruited using the online survey platform Prolific completed the survey (129 female, 71 male; aged 18 to 71, M=34.7, SD=12.1; 17.5% were students; 50% in full time employment, 21.5% in part time employment; 28.5% unemployed or other) and were paid £0.70 for their time. The sample was calculated to have a power of 0.8 to detect a minimum effect size of d=0.39. Participants were selected on the basis of British nationality and residency in England and Wales. The study was conducted in accordance with approval obtained from UCL's Research Ethics Committee (EP/2018/005). Informed consent was obtained from each individual in advance of participation by providing them with study information and a consent form to agree to.

### 5.3.1.2 Design

We used a 2 × 2 mixed design in which all participants viewed both cases, the apprentice scenario and the captain scenario, sequentially but in a randomised order. As with Study 8, as all participants viewed both scenarios, scenario was a within subject manipulation, whereas participants only viewed the scenarios in one of two possible orders, hence order was a between subject manipulation.

### 5.3.1.3 Materials

Participants were advised that they would be put in the place of a juror in a crown court and asked to consider two hypothetical cases where the accused was relying on the defence of duress of circumstances and that they would be asked to determine whether the accused should be convicted or acquitted. It was explained to participants that the defence of duress of circumstances is where an accused commits what would otherwise be an offence in order to save a life or prevent serious harm to somebody.

In both scenarios it was explained that a small 3-man crew were employed to transport a new yacht to Australia and that part-way through the voyage and when a long distance from land, the yacht was fatally damaged in a storm and rapidly sank. The crew escaped to the lifeboat, but without communication equipment. Participants were told in both cases that over a period of 2 weeks, the crew exhausted the lifeboat's rations, they were unable to catch fish or seabirds, and only collected a tiny amount of water. After a further week with no provisions, the situation became desperate. The crew had seen no other ships in the 3 weeks, estimated that they were around 1,000 miles from land, and that there seemed no immediate prospect of rescue. Thereafter the scenarios diverged. In the apprentice scenario, participants were told that the captain and mate secretly discussed the situation and decided that they might survive for a little longer if they killed and ate the apprentice, who they felt would be likely to die soonest. Participants were advised that after the apprentice fell asleep, the mate held him down while the captain killed him with a knife. Over the subsequent days the captain and the mate consumed the apprentice. This was relatively similar to the original case. By contrast, in the captain scenario, participants were advised that the 3 crew members discussed the situation together and decided to draw lots to decide who would be sacrificed. The captain drew the short straw and allowed the mate and apprentice to kill him with a knife. Participants were told that over the subsequent days, the mate and apprentice consumed the captain. Following that, participants were told in both scenarios that 6 days later a ship was seen and the remaining sailors used their final distress flare successfully to attract its attention leading to their rescue. The rescuers witnessed the remains of the sailor who had been killed.

In both scenarios, participants were given further details of the legal issues, in particular the defence of duress of circumstances. This included the information that (1) duress of circumstances is where the accused committed the offence because they reasonably believed that they would die or be seriously injured and (2) a sober person of reasonable firmness, sharing the characteristics of the accused, would have acted in the same way. Because the scenario was this time a criminal case, it was explained that the burden of proof was on the prosecution. Thus participants were advised that if they were sure that one or both of those statements was untrue, they should find the accused guilty and if they thought that both of those statements was or may be true, they should find the accused not guilty.

Participants were then asked how they found the accused, guilty or not guilty, to indicate on a scale from 0-100 how reasonable it was for the accused to argue the defence of duress of circumstances, and they were also asked to explain their decision.

#### 5.3.1.4 Measures

As with the previous experiment, participants were asked to respond to three measures. The first measure was simply a verdict on the accused, as would be expected in a criminal trial. This was a binary choice between guilty or not guilty (for both accused). Similar to the previous experiment, in order to glean a more nuanced view of participants' views as well as for consistency with previous moral psychology trolley-type research, participants were also asked to give a view of the perceived reasonableness of the accused arguing duress of circumstances. This was a 100-point scale from 0 (representing completely unreasonable) to 100 (representing completely reasonable) with 1-unit gradations. Finally, as with the earlier experiment, participants were also asked to explain their decision and were provided with an open text response field. While this survey was based on a criminal jury trial in which the jury would not be expected to give reasons, individual jurors in a criminal trial would be expected to share their reasons with other jurors as part of the group deliberations, so the request for an explanation seemed reasonable.

After participants had reviewed the scenarios, they were also asked some further questions. One question was whether they thought that their response to the first scenario affected their response to the second scenario, with participants permitted to select from a ternary of 'yes', 'no', and 'don't know'. Participants were then asked to indicate a binary response to which statement best reflects their view: 'similar legal cases should be treated the same' or 'each legal case should be decided on its own merits. Participants were also asked as an attention check to indicate what they were asked to decide given a choice of three statements of which they could choose as many options as they wished: 'whether the accused should be acquitted or found guilty', 'whether the accused should be allowed to argue defence of circumstances', and 'whether the case should be referred to the Court of Appeal'. Participants were finally asked whether they found the explanation of the law easy or difficult to understand and to explain if they had any relevant legal experience or knowledge.

#### 5.3.1.5 Procedure

As noted above, participants were recruited online and participated in the survey using the online platform Qualtrics in a place of their choosing, using their own device. On initial referral from the Prolific platform, participants were first provided with the study information form. They were then asked to complete the consent form comprising a number of statements to which an affirmative answer was required in order to participate in the survey. An anonymous user identification was collected to enable subsequent matching of demographic data without compromising the participants' anonymity.

Participants were randomly assigned to one of the two conditions by the Qualtrics platform such that half of the participants viewed the apprentice scenario followed by the captain scenario and half the participants viewed the captain scenario followed by the apprentice scenario. After reviewing the first scenario that they were allocated to, they were first asked whether they would find the accused guilty or not guilty, secondly how reasonable

they found the defence on a scale of 0 to 100, and thirdly to explain their decision. Once they had reviewed and responded to the measures on the first scenario, they were then presented with the second scenario. After completing the responses to the scenarios, they were posed the additional measures referred to above.

After completing the survey, participants were thanked for their participation and referred back to the Prolific survey platform to confirm their participation. Once both platforms had confirmed the participant's successful completion of the survey, their remuneration was authorised.

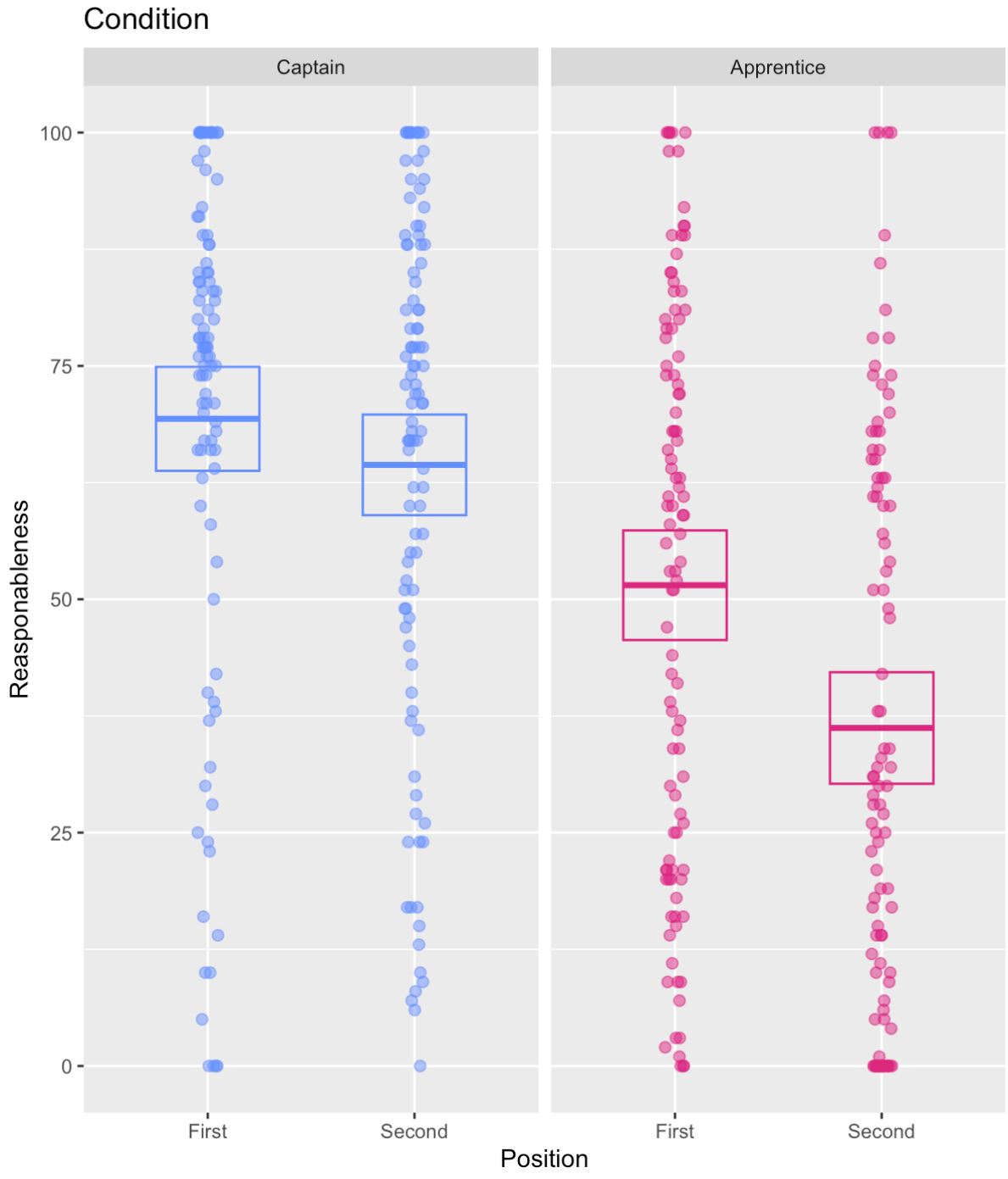
### 5.3.2 Results

Most participants (79%) found the explanation of the law easy to understand, with a small proportion (21%) finding it difficult. For the attention check, all participants indicated that the issues in the case were whether the accused should be acquitted or found guilty or whether the accused should be allowed to argue defence of circumstances or both, with no participants indicating the clearly erroneous option of whether the case should be referred to the Court of Appeal. A small proportion (5%) had some modest legal familiarity such as having studied a law degree, but only a handful (1%) had heard of the case of *R v Dudley v Stephens* (1884) 14 QBD 273 DC. As such, the decision was taken not to exclude any participants.

In terms of responses to the scenarios considered in isolation, results were consistent with responses received during pre-testing. Participants were much more likely to find the accused guilty in the apprentice scenario (73:30 guilty:not guilty), but this pattern was reversed in the captain scenario (37:62 guilty:not guilty). A Fisher's exact test confirmed that this pattern was statistically significant ($p<0.001$, OR=4.05, 95% CI [2.17, 7.67]). Similarly, the mean reasonableness assessment in the apprentice scenario was lower at 51.5/100 than that in the captain scenario of 69.3/100. This difference was statistically significant

(t(200)=4.36, 95% CI = [9.78, 25.9], Cohen's D= 0.61, p<0.001), see Figure 12.

*Figure 12. Reasonableness responses by condition and position for Study 9.*




However, while an order effect of the scenarios was observed, and the apprentice scenario was labile while the captain scenario was stabile, the most interesting finding was

that the order effect was in the opposite direction to that predicted, with participants more likely to convict and viewing the accused as less reasonable in the apprentice scenario when they reviewed it after the (more acceptable) captain scenario compared to reviewing it in isolation, see Figure 12. Thus, responses in the captain scenario remained fairly sympathetic to the accused even when viewed after the apprentice scenario. Participants preferences for a guilty verdict in the captain scenario increased slightly after viewing the apprentice scenario (from 37:62 guilty:not guilty to 49:54), but this difference did not reach statistically significant according to a Fisher's exact test ($p=0.16$, $OR=0.66$, 95% CI = [0.36, 1.20]). Similarly, mean reasonableness assessments remained high in the captain scenario whether viewed in isolation (69.3/100) and only dropped slightly when viewed after the apprentice scenario (64.4/100). Likewise, this difference was not statistically significant ($t(200)=1.26$, 95% CI = [-2.77, 12.64], Cohen's $D= 0.18$, $p=0.21$). More notably, contrary to predictions, responses in the apprentice scenario were labile, but became less similar to the captain scenario that preceded it rather than more similar. Thus guilty verdicts increased from 73:30 (guilty:not guilty) to 81:18, a difference that was borderline statistically significant according to a Fisher's exact test ($p=0.07$, $OR=0.54$, 95% CI = [0.26, 1.10]). In line with this, assessments of reasonableness decreased in the apprentice scenario between when it considered in isolation (51.5/100) and when it was considered after the more acceptable captain scenario (36.2/100). This difference was statistically significant ($t(200)=3.62$, 95% CI = [6.97, 23.62], Cohen's $D= 0.51$, $p<0.001$).

In terms of the principles that participants were willing to endorse, most participants overall (88%) thought that each legal case should be considered on its own merits and only a minority (12%) endorsed the principle that similar legal cases should be treated the same. Despite this, a majority (63%) of participants in fact gave a consistent verdict across the two scenarios with only a minority (37%) giving inconsistent verdicts. Notably, participants who gave consistent verdicts were more likely to endorse consistency as a principle (16%) compared to those who gave inconsistent verdicts (5%). This last difference was statistically significant based on a Fisher's exact test ($p=0.03$, $OR=0.29$, 95% CI = [0.07, 0.92]), consistent with either participant's values influencing their verdicts, or participant's verdicts

influencing the values they were prepared to endorse publicly.

Concerning participants' subjective insight into whether they felt they had been influenced by the previous case, most (61%) said that they did not feel that they had been influenced. A corresponding minority (39%) felt that they had been influenced. Those participants who felt that they were influenced by the prior case were more likely to give a different verdict between scenarios (45%) than those who felt that they were not influenced who were less likely to give a different verdict between the scenarios (31%). This difference was borderline statistically significant at the 0.05 level according to a Fisher's exact test ($p=0.09$, $OR=1.76$, 95% $CI=[0.92, 3.38]$). Given the unexpected finding above that the order influence appeared to be that the captain scenario (in which participants were more likely to acquit) made participants subsequently seeing the apprentice scenario more likely to convict, the evidence seemed somewhat consistent with participants' views.

### 5.3.3 Discussion

Results from Study 9 revealed a number of important findings. The headline finding is that the research paradigm used seemed to confirm the existence of further order effects in paired dilemmas in a new context, that of legal decision-making and in particular in a criminal law - rather than civil law - context. Consistent with previous findings, the order effect appeared asymmetric with one dilemma seemingly being influenced by the other, but not vice versa. In the captain scenario, where the starving sailors killed and ate the most senior sailor on board following a fair selection process, was generally more acceptable to participants. Thus participants rated the sailors' behaviour as more reasonable and were more likely to accept the defence of duress of circumstances, leading them to prefer acquittal. This pattern remained stabile regardless of whether the captain scenario was considered first (and thus in isolation) or whether it was considered after the generally less acceptable apprentice scenario. By contrast, in the apprentice scenario, where the starving sailors killed and ate the most junior sailor on board following an unfair selection process, participants tended to find

this much less acceptable. Correspondingly, participants rated the sailors' behaviour as less reasonable and were less likely to accept the defence of duress of circumstances, leading them to prefer conviction. However, the pattern of responses was much more labile in the apprentice scenario, with participants finding the sailors' behaviour less reasonable and convicting more when they viewed this scenario after the captain scenario, compared to when they saw the apprentice scenario in isolation.

What was particularly notable was the unexpected direction of the order effect. Previously reported results of order effects in paired dilemmas, primarily from the much-studied moral decision-making field invariably indicated that the order effects caused responses in similar dilemmas to become more similar to one another. In this study we found the influence had the opposite effect: responses in the labile dilemma became more dissimilar to responses in the dilemma that preceded it. While this is currently an isolated finding, it is potentially quite consequential, in particular because it gives more credence to one theoretical finding while casting doubt on another. In terms of the two theories most pertinent to these phenomena, the finding does not seem to be easily reconcilable with consistency or coherence type theories in that these theories posit that participants are trying to be consistent across scenarios, predicting that responses should become more rather than less similar. While it seems likely that there are some situations where participants try to appear consistent when considering similar dilemmas, particularly where inconsistency reveals that they have taken impermissible factors into account when reaching their decision (Nadler, 2012, p. 26; Sood & Darley, 2012, pp. 1343–1344), it does not seem likely that this is the most important influence in these circumstances. By contrast, the findings do seem to accord with salience type theories that suggest that order effects may be caused by earlier scenarios drawing participants' attention to aspects of the situations that they knew of but had not realised the salience of. Obviously salient information could affect responses in either direction. Nonetheless, many questions remain unanswered, such as precisely what about the different information presented in the captain scenario might be salient. One might speculate that the behaviour of the sailors in the apprentice scenario might appear quite reasonable on a relatively superficial examination, but that the presentation of the captain scenario might

cause participants to focus on salient information that they had not previously considered, such as that a fairer way of selecting the individual to be cannibalised was possible that did not lead to the death of the most vulnerable sailor.

There are obviously other differences between the earlier Study 8 and Study 9, not least that it was a criminal context rather than a civil context, that participants this time were deciding on the reasonableness of behaviour that had already been undertaken by a third party in the past rather than making a decision themselves that would have consequences in the future, but it is difficult to see a theoretical reason why these factors would have made such a significant difference.

One issue with the previous study was that participants were laypeople when the type of legal problem posed was one that would invariably be taken by a relatively senior judge. The topic of the present study addressed this fact to some extent in that the type of decision taken was one that would ordinarily be taken by laypeople and the participants chosen would have been eligible, in principle, to act as jurors. There were nonetheless issues with the validity of the experiment, most prominently the fact that there was no group deliberation dimension to the participants' task and that ordinarily jurors are not expected to give reasons for their group decision. Nonetheless, as real-life jurors would necessarily give reasons for their thinking to other jurors during the deliberation phase, the requirement to give reasons was consistent with this, even if there was no equivalent of deliberation. Nonetheless, the paradigm was useful for revealing the phenomenon, particularly given the considerable complexity and cost of carrying out jury research. Once the phenomenon is better understood in these more constrained circumstances, replication in a more plausible jury context would be advisable.

While the findings of the current study are significant and novel, some care needs to be taken before extrapolating more theoretical implications from them. This was a new paradigm in a context where there have been relatively few attempts to isolate order effects through the presentation of similar pairs of legal dilemmas. As such, before undertaking

research that addressed the previously identified imperfections in the paradigm, replication would be important.

## 5.4 STUDY 10

In order to clarify the findings of Study 9, a related criminal scenario was chosen for Study 10 that would permit us to focus on one of the two differences between the scenarios, namely the method of selection rather than the vulnerability of the victim. This scenario was based on another incident at sea that slightly pre-dated the case of *R v Dudley v Stephens* and that also resulted in a criminal prosecution: *US v Holmes* 1 Wall Jr 1 (1842). This trial, prosecuted in the United States, resulted from the wrecking of the ship the William Brown. In the real-life case, the ship was sunk after hitting an iceberg and the crew made it to two boats: a jolly boat and a long boat. The first mate, Francis Rhodes made it to the long boat together with several crew and passengers. Nonetheless, the boat was leaky and dangerously overloaded. Of the sailors on the long boat, Rhodes was the most senior. He took the decision to throw passengers overboard to prevent the overloading. Sailors, including one called John Holmes, threw a number of male passengers overboard. Given the frigid temperatures, all perished. However, the long boat was subsequently spotted and the crew and passengers were rescued. When the passengers finally reached their destination in Philadelphia, they complained to the authorities. Holmes was the only sailor known to be in Philadelphia, and he was prosecuted for the manslaughter of one of the passengers who had been thrown overboard, convicted, and sentenced to six months' imprisonment and a fine. *Holmes* was chosen as the basis for the study for being similar to *Dudley v Stephens* in that those to leave the boat were chosen unilaterally by those responsible rather than being selected at random. It was also a much less familiar legal case for those from the jurisdiction of England and Wales.

As with the previous studies, two versions of the scenario were prepared based on the *Holmes* case. Whereas Study 9 had two differences between the scenarios (method of selection and vulnerability of the victim), this study was pared down to simply the method of

the selection. Thus in one scenario the victims were chosen unilaterally by the captain, whereas in the other scenario, the victims were chosen following a fair procedure. In the 'unilateral' scenario, the crew and passengers of a wrecked yacht were floating in a dangerously overloaded life-raft. The captain's response was to unilaterally choose male passengers to be forcibly removed the raft. In the 'lottery' scenario, lots were drawn from among the male crew and passengers. However, after lots had been drawn, one of the passengers who had drawn the short straw refused to leave voluntarily, so the captain and crew forcibly removed him from the raft. In both scenarios the captain was prosecuted for manslaughter. As before, there was no dispute on the facts and the captain's argued duress of circumstances to avoid liability. Pretesting confirmed that participants viewed the unilateral scenario as worse (mean reasonableness = 3.85/7 n=10) than the lottery scenario (mean reasonableness = 4.5/7 n=10).

Although the manipulation in this study appeared to be more limited than in the previous study in that the focus was limited to the means of selection, we expected there to be a similar pattern with a similar, or smaller, effect size. In particular, we predicted that the lottery scenario would remain stabile, but the unilateral scenario to be labile, with participants finding the unilateral scenario less reasonable after then had reviewed the lottery scenario compared to when they reviewed the unilateral scenario in isolation.

## 5.4.1 Method

### 5.4.1.1 Participants

Two hundred participants were recruited using the online survey platform Prolific on the basis of British nationality and residency in England and Wales. Of these, one hundred and seventy five satisfactorily completed the attention check. Of these, 108 (62%) were female and 67 (38%) were male; ages were from 18 to 71, M=35.5, SD=13.0; 21% were

students; and 48.0% were in full-time employment, 19.4% were in part-time employment, and 32.6% were unemployed or of other status. The study was calculated to have an 80% power to detect an effect size of d=0.40. Participants were paid £0.60 for their time. The study was conducted in accordance with approval obtained from UCL's Research Ethics Committee (EP/2018/005). Informed consent was obtained from each individual in advance of participation by providing them with study information and a consent form to agree to.

#### 5.4.1.2 Design

We used a 2 × 2 mixed design in which all participants viewed both cases, the unilateral scenario and the lottery scenario, sequentially but in a randomised order.

#### 5.4.1.3 Materials

As before, participants were advised that they would be put in the place of a juror in a crown court and asked to consider two hypothetical cases where the accused was relying on the defence of duress of circumstances and that they would be asked to determine whether the accused should be convicted or acquitted. Again, it was explained to participants that the defence of duress of circumstances is where an accused commits what would otherwise be an offence in order to save a life or prevent serious harm to somebody.

In accordance with the established paradigm, participants considered two similar scenarios sequentially. In both scenarios, participants were told that a charter yacht carrying passengers had been caught in a severe storm and rapidly sank, taking most of the life rafts with it. Both scenarios described the captain, 5 crew, and 32 passengers making it to the one remaining life raft, a life raft that was only designed for 15 people and dangerously overloaded. Water was lapping at the door and washing inside. By constant bailing, the crew and passengers were able to maintain its level in the water. However, the captain and crew

thought that even a moderate blow from a wave would swamp the raft causing it to sink. Night was falling and any prospect of rescue seemed unlikely until the following day. The captain and crew thought that given the weather conditions and unless the weight was reduced, the raft would sink during the night with the inevitable loss of most, if not all, on board.

In the lottery scenario, participants were advised that the captain discussed the situation with those on board and everybody agreed that lots would be drawn from the able-bodied crew and passengers to decide who would leave the raft. Those who left the raft would have to hang onto the outside of the raft while floating in the sea. Lots were drawn and most of those selected voluntarily left the raft. However, given the unlikely prospects of survival outside the raft, some passengers who had been selected by lot then refused to leave. The captain, assisted by 2 crew members, forcibly removed those passengers from the raft. The unilateral scenario was the same as the lottery scenario, save that participants were instead told that the captain unilaterally decided that some able-bodied passengers and crew would be required to leave the raft. When some selected passengers refused to leave, the captain assisted by two crew, forcibly removed the passengers from the raft.

Both scenarios were thereafter the same. Participants were advised in each scenario that the raft stayed afloat until the following day when the remaining passengers and crew were rescued. However, those who had left or removed from the raft were lost. Participants were advised in both scenarios that the captain was being prosecuted for manslaughter for forcibly removing passengers from the raft and that he was relying on the defence of duress of circumstances. Participants were told that that defence is where the accused committed the offence because they believed that otherwise death would result and a person of reasonable firmness would have acted in the same way. Participants were also told that the prosecution accept that the captain may have believed that the lives of the passengers and crew were at risk, but that he was not acting reasonably by forcibly removing the passengers from the raft who refused to leave.

This time, passengers were asked for each scenario to indicate their responses to a single sliding 7-point Likert scale with verbal descriptions corresponding with each point. Passengers were also asked to explain their decision.

#### 5.4.1.4 Measures

In this study, participants were posed a single main measure in response to each scenario. As noted above, this was a 7-point Likert scale with verbal descriptions corresponding to each of the 7 points. These were: 1 = completely unreasonable / 2 = unreasonable / 3 = somewhat unreasonable / 4 = evenly balanced / 5 = somewhat reasonable / 6 = reasonable / 7 = completely reasonable. Participants could indicate anywhere along that scale in gradations of 0.1 from 0 to 7. Participants were also required to give an explanation of their decision in an open-ended text field.

Once participants had responded to the two scenarios, they were asked additional questions. These included an attention check asking what they had been asked to decide which listed two options, how reasonable it was for the accused to forcibly remove passengers from the raft and whether the captain was negligent, with participants able to choose either or both of these options. Participants were also asked whether they thought their response to the first scenario affected their response to the second scenario and which statement best represented their view: that similar cases should be treated the same or that each legal case should be decided on its own merits. Participants were also asked to indicate whether they found the explanation of the law easy or difficult to understand.

#### 5.4.1.5 Procedure

As with previous experiments, participants were recruited online from the Prolific

platform and participated in the survey using the online platform Qualtrics in a place of their choosing, using their own device. On initial referral , participants were first given with the study information form. They were then asked to complete the consent form comprising a number of statements to which an affirmative answer was required in order to participate in the survey. An anonymous user identification was collected to enable subsequent matching of demographic data without compromising the participants' anonymity.

Participants were randomly assigned to one of the two conditions by the Qualtrics platform such that half of the participants viewed the lottery scenario followed by the unilateral scenario and half the participants viewed the unilateral scenario followed by the lottery scenario. After reviewing the first scenario that they were allocated to, they were first asked how reasonable it was for the captain to forcibly remove the passengers from the raft when they refused to leave, indicating their responses on the 7-point Likert scale previously described. They were then asked to explain their decision. Once they had reviewed and responded to the measures on the first scenario, they were then presented with the second scenario and posed the same measures. Only after they had responded to both scenarios were they asked to respond to the additional measures referred to above.

After completing the survey, participants were thanked for their participation and referred back to the Prolific survey platform to confirm their participation. Once both platforms had confirmed the participant's successful completion of the survey, their remuneration was authorised.

### 5.4.2 Results

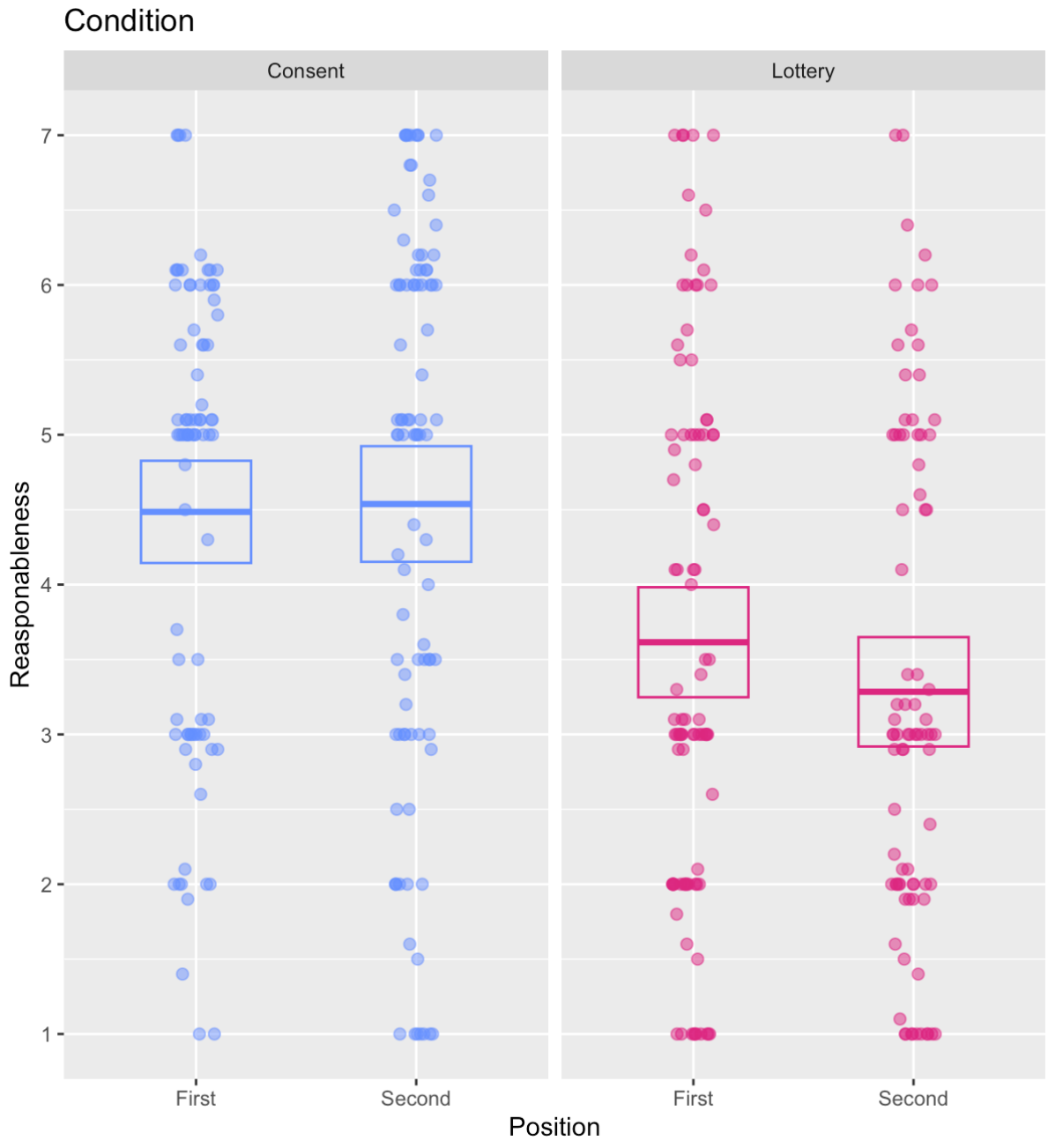
Of the 200 participants who successfully completed the survey, 25 incorrectly answered that one of the issues they had been asked to determine was whether the captain was negligent. These participants were excluded from the analysis. Of the 175 remaining participants, 89% found the explanation of the law easy to understand with only 11% finding

it difficult.

Considering participants responses to the two scenarios in isolation, the results were consistent with pretesting. Participants found the captain's response in the lottery scenario much more reasonable (M=4.49/7, SD=1.56) than in the unilateral scenario (M=3.62/7, SD=1.77). A two-sample t-test confirmed that this difference was significant (t(173)=3.41, 95% CI = [0.37, 1.37], p<0.001, Cohen's D=0.52).

As expected, the lottery scenario changed very little regardless of whether participants reviewed it in isolation or after the unilateral scenario, see Figure 13. Mean assessments of reasonableness remained relatively favourable regardless of whether the scenario was assessed in isolation (M=4.49/7, SD=1.56) or whether it was assessed after the unilateral scenario (M=4.54/7, SD=1.86). According to a two-sample t-test, this difference was not significant (t(173)=0.20, 95% CI = [-0.57, 0.46], p=0.84, Cohen's D=0.03].

*Figure 13. Participants' reasonableness assessments for the consent and lottery scenarios by order of presentation in Study 10.*




The unilateral scenario changed in the direction predicted, but the change did not reach statistical significance, see Figure 13. While initially less favourable (M=3.62/7,

SD=1.77) than the lottery scenario when presented in isolation, it became even less favourable (M=3.28/7, SD=1.67) when reviewed after the lottery scenario. According to a two-sample t-test, this difference was not significant (t(173)=1.27, 95% CI = [-0.18, 0.85], p=0.21, Cohen's D=0.19).

Consistent with the primary findings above, there was the greatest difference between the means of the reasonableness ratings given by participants when they saw the scenarios in the order lottery>unilateral (1.20) compared with when they saw the scenarios in the order unilateral>lottery (0.92). This also suggested that the unilateral was likely to be the more labile scenario and that participants found it less reasonable when reviewed after lottery. In accordance with this, in response to the question whether the first scenario had influenced their response to the second scenario, slightly fewer thought that they had been influenced (45%) than thought that they had not been influenced (55%). In line with the greater difference in responses for participants who had seen the scenarios lottery>unilateral, these participants were also more likely to report that they had been influenced by the prior scenario (48%) than those seeing the scenarios in the order unilateral>lottery (42%). Nonetheless, a Fisher's exact test indicated that this difference was also not statistically significant (p=0.52, OR=1.26, 95% CI= [0.64, 2.49]). This time there was no difference between the statements that participants were willing to endorse, regardless of which order they reviewed the scenarios. Overall, most participants preferred to endorse the statement that 'each legal case should be decided on its own merits' (87%) rather than 'similar legal cases should be treated the same' (13%), but there was effectively no difference between the proportions in each condition.

### 5.4.3 Discussion

Study 10 was similar to Study 9 in a number of respects, yet the results were not statistically significant. Similarities between the cases included the dilemma of whether the commission of a serious criminal wrong was justified by the necessity to avoid a greater

harm; the balancing of different numbers of lives saved or lost; the means of selection of those to be sacrificed; and the isolated maritime context which limits other possible courses of action. The key difference was in the victims. In Study 9 the victim was either the most senior or the most junior sailor, whereas in Study 10 the victims were always the passengers. Other differences included the nature of the threat: either the risk of starvation in due course or the risk of being swamped and sunk by a wave; but these did not seem to be likely to be material. Equally, the effect seen in Study 9 may have either been an artefact of the measures used or the lack of effect identified in Study 10 might have been caused by the different measures in the latter study, but this did not seem very likely. Focussing on the means of selection, it seems obvious that an objectively fair process is preferable to an individual unilaterally choosing who to be sacrificed. While from one perspective, a sailor picking passengers could be thought of as almost as random as a lottery, a risk would exist that subjective considerations of the kind explored in Sections 2, 3, and 4 would influence the choice, thereby making a lottery somewhat preferable. The fact that there was a statistically significant difference between how reasonable participants assessed the scenarios seemed to bear this out.

In relation to the hypothesised order effect of the unilateral scenario being assessed as less reasonable when reviewed after the lottery scenario, while the difference was not statistically significant, it was in the direction predicted. Given this was on a 2-tailed test assessing a simple difference between the reasonableness ratings, this is some evidence for an effect running in the opposite direction of consistency. Overall, the evidence could be consistent with either no effect or with the sample being insufficiently powerful to detect the effect of interest. As noted at the outset, it was envisaged that any effect would be equivalent or smaller than that seen in Study 9 on the basis that the manipulation in the later study was deliberately more limited than in the earlier study. In particular, in Study 10, the selection was limited to some or other passengers. While there was the potential for subjectivity to creep into the selection process, there was not the same risk of the most powerful individuals taking advantage of the most vulnerable as there was in the earlier study.

In addition to similar questions about the external validity of Study 10 as with Study 9, the introduction of the alternative measure entailed some compromises. In Study 9 we used two different primary measures to assess participants' responses to the scenarios. One of these was a binary response that was more externally valid as it reflected the binary decision that a real-world juror would be required to make whereas the other was a more sensitive gradated measure more akin to the measures used in the moral decision-making context. Reducing these to a more user-friendly single Likert scale with verbal descriptions in Study 10 should have served to facilitate participants' responses, but could also have undermined the external validity of the study. While this seems to be a relatively small risk, it would be wise to validate the measure using the paradigm and effect seen in Study 9.

The equivocal findings of Study 10 suggest avenues for further research. In particular, given the unknown size of any effect, it would be advisable to conduct a much higher-powered study to determine what, if any, effect exists.

## 5.5 STUDY 11

Given the equivocal results of Study 10 in the context of the new paradigm, for Study 11 we decided to revert to the original paradigm to see if the findings could be replicated. At the same time, we took the opportunity to look for further illumination into the phenomenon previously observed. Because the most plausible explanation seemed to be that the order effect seen in the unilateral scenario was caused by participants realising (through actively completing the lottery scenario) that there was a better way of addressing the dilemma, we sought to bring this about using a different means. The most obvious means to do this seemed to be simply to disclose the information to one group of participants before they made their assessment.

This time, we adopted a different experimental design in which all participants saw the same version of the previous unilateral scenario from Study 9. In order to replicate the effect seen in that study, a control group saw only the unilateral scenario whereas another

replication group assessed the lottery scenario prior to this unilateral scenario. In order to try to bring about the effect using a different manipulation, a third group assessed the unilateral scenario after hearing submissions by putative prosecution and defence advocates. In particular, in that condition, the prosecutor specifically raised the point that the behaviour of the sailors was not reasonable because there were other, more reasonable, options available to them. The prosecutor specifically gave the example of randomly drawing lots as a more reasonable means of behaving. To try to avoid the potential confounding effects of persuasion, the prosecutor's submissions were neutral and descriptive, and the defence submissions were also concise and banal.

In terms of the replication of the effect previously seen, we predicted that participants in the group who undertook the lottery scenario prior to the target unilateral scenario would assess the lottery scenario as less reasonable than those in the control group who assessed the unilateral target scenario in isolation. In addition, we predicted that those in the group who were exposed to the prosecutor's submissions disclosing that there was a better means to select those to die would find the lottery scenario less reasonable than those in the control group where there were no submissions.

## 5.5.1 Method

### 5.5.1.1 Participants

Three hundred participants were recruited using the online survey platform Prolific on the basis of British nationality and residency in England and Wales and successfully completed the survey. Of these, 187 (62%) were female and 113 (38%) were male; ages were from 18 to 72, M=36.3, SD=12.2; 16% were students; and 52% were in full-time employment, 21% in part-time employment, and 27% were unemployed or of other status. The study was calculated to have an 80% power to detect an effect size of $\eta 2 = 0.038$.

Participants were paid £0.50 for their time. The study was conducted in accordance with approval obtained from UCL's Research Ethics Committee (EP/2018/005). Informed consent was obtained from each individual in advance of participation by providing them with study information and a consent form to agree to.

#### 5.5.1.2 Design

We used a between participant design with one independent variable with three levels in which all participants viewed the target unilateral scenario. Participants were randomly assigned to 3 conditions: a control condition where participants reviewed only the unilateral scenario; a replication condition where participants reviewed the lottery scenario before the unilateral scenario; and a new condition where participants additionally read closing submissions that highlighted the existence of an alternative solution for choosing who to be killed (namely a lottery), before making their decisions.

#### 5.5.1.3 Materials

All participants were advised at the outset that they would be asked to put themselves in the place of a juror in the Crown Court. Duress of circumstances as explained as a defence that was available where an accused commits what would otherwise be an offence in order to save a life or prevent serious harm to somebody. Participants were told that more details would be provided with the cases, and that they would be asked to decide one or two hypothetical cases where the accused argues defence of circumstances, and asked to give a verdict.

In line with the materials from Study 9, all participants were asked to consider the unilateral scenario. They were advised that while a very great distance from land, the yacht

was fatally damaged in a storm and rapidly sank, with the three crew escaping to the lifeboat, but without communication equipment. Over 2 weeks the exhausted the lifeboat's rations, were unable to catch fish or seabirds for food, and collected only a tiny amount of rainwater. They used most of the distress flares without success. A further week later, the situation was desperate, they had seen no other ships for 3 weeks and they estimated they were around 1,000 miles from land. The captain and the mate secretly discussed the situation and decided that they might survive a little longer if they resorted to cannibalism, and they agreed to kill the apprentice. After the apprentice fell asleep, the mate held him down while the captain killed him with a knife, and they consumed him over the subsequent days. 6 days later, a ship was seen, and the captain and mate used the final distress flare to attract attention. The rescuers witnessed the remains of the apprentice and the evidence of cannibalism. Participants were told that the captain and mate were being prosecuted for murder, that the defence accepted that they committed what would ordinarily be murder, but relied on the defence of duress of circumstances before the jury. Duress of circumstances was explained as where the accused committed the offence because they reasonably believed that otherwise they would die or be seriously injured and a person of reasonable firmness would have acted in the same way. The legal issues were simplified slightly from Study 9. Participants were told that the prosecution accepted that the captain and mate may have believed that their lives were at risk, but argued that they were not acting reasonably by killing the apprentice. It was finally explained that because the burden of proof was on the prosecution: (1) if they were sure the accused were acting unreasonably by killing the apprentice, they should find them guilty; and (2) if they thought that the accused were or might have been acting reasonably by killing the apprentice, they should find them not guilty. Participants were asked to give a verdict, to assess the reasonableness of the accused's actions, and to give reasons for their verdict.

For those participants assigned to the control condition, the unilateral scenario was the only scenario they were asked to assess. For participants assigned to the replication condition, they were asked to assess a version of the lottery scenario prior to the unilateral scenario. This was the same as the unilateral scenario other than the fact that all of the crew discussed

the situation and decided to draw lots to decide who would be sacrificed. In this scenario, the captain drew the short straw, allowing the mate and the apprentice to kill him with a knife. As before, the surviving members consumed the crew member who had been killed. In the new condition, participants viewed a variant of the unilateral scenario that included submissions by both the prosecution and the defence prior to making their decision. The prosecution submissions specifically drew attention to the assertion that the behaviour of the accused was not reasonable because there was a better way to behave, namely to randomly draw lots to decide who would be sacrificed.

After completing the relevant materials appropriate to each condition, participants completed an attention question in which they were asked to identify the issues from a selection of three: (1) whether the accused should be acquitted or found guilty, (2) how reasonable it was for the accused to kill when faced with starvation, and (3) whether the accused's lives were at risk. Of these, the third was the erroneous answer given that participants were advised that the prosecution accepted that the accused may have believed that their lives were at risk. As before, participants were asked to indicate whether they found the description of the law easy or difficult to understand, and whether they had any previous legal experience.

#### 5.5.1.4 Measures

For each scenario, participants had to respond to the same measures. These were a binary indication of verdict which was either guilty or not guilty; a gradated Likert response with 0.1 increments from 1-7 where each number was matched with a verbal description (1 = completely unreasonable / 2 = unreasonable / 3 = somewhat unreasonable / 4 = evenly balanced / 5 = somewhat reasonable / 6 = reasonable / 7 = completely reasonable); and a text response box to give an explanation of their decision.

Once participants had responded to one scenario or both, all participants responded to

a series of other measures. An attention check invited the participants to identify the issues in the case with 2 correct answers (whether the accused should be acquitted or found guilty, how reasonable it was for the accused to kill when faced with starvation) and one incorrect answer (whether the accused's lives were at risk). Participants were also asked if they found the explanation of the law easy or difficult, if they had any relevant legal knowledge or experience, and if they found any parts of the survey confusing or inconsistent.

#### 5.5.1.5 Procedure

Again, participants were recruited online from the Prolific platform and participated in the survey using the online platform Qualtrics in a place of their choosing, using their own device. On initial referral, participants were first given with the study information form. They were then asked to complete the consent form comprising a number of statements to which an affirmative answer was required in order to participate in the survey. An anonymous user identification was collected to enable subsequent matching of demographic data without compromising the participants' anonymity.

Participants were randomly assigned to one of the three conditions by the Qualtrics platform such that a third of the participants viewed the unilateral scenario only; a third viewed the lottery scenario before the unilateral scenario; and a third viewed a version of the unilateral scenario in which the prosecutor submitted that the accused were unreasonable because they could have drawn lots. After viewing a scenario, they were first asked to give a verdict; second asked to assess the reasonableness of the accused's action on the Likert scale; and third asked to explain their decision. Once they had completed the one scenario or two, all participants were then posed the additional measures described above.

After completing the survey, participants were thanked for their participation and referred back to the Prolific survey platform to confirm their participation. Once both platforms had confirmed the participant's successful completion of the survey, their

remuneration was authorised.

### 5.5.2 Results

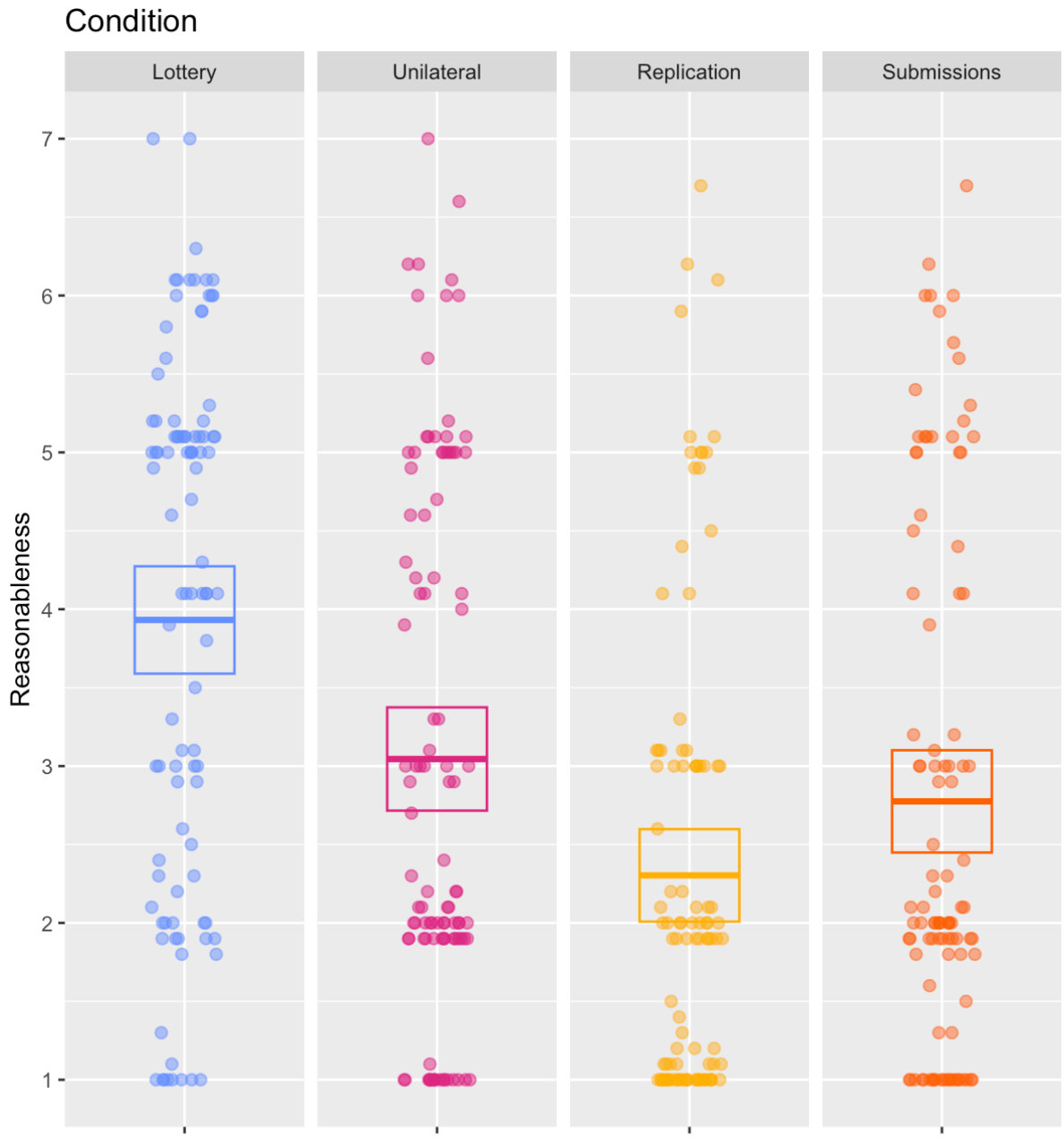
300 participants completed the survey. In terms of the attention check, it was assessed that it may have been too challenging for participants given that the 'incorrect' answer required participants to appreciate that the issue of whether the accused lives were at risk was only not a live issue because the prosecution in this version of the survey had conceded it. Nonetheless, many participants were sensitive to this, with only 9% of participants choosing this response. Relatedly, the explanation of the law in this survey seemed to be better understood by participants, with a lower percentage (9%) compared to Study 2 finding the explanation of the law difficult to understand. Again, only a small percentage (7%) had any legal experience, though for the most part this was limited, no reference was made to any familiarity with the case of *R v Dudley v Stephens*, and this would ordinarily not have been sufficient to exclude such participants from acting as jurors in such a case. As a result, the decision was taken not to exclude any participants.

Although the lottery scenario was a manipulation rather than a condition, the incidental data from participants confirmed once again that the lottery scenario was assessed as more favourable (reasonableness M=3.93/7, SD=1.71, verdicts guilty:not guilty = 42:56) than the unilateral scenario (reasonableness M=3.05/7, SD=1.68, verdicts guilty:not guilty = 72:30). See Figure 14, first panel. These differences were significant according to a t-test and a Fisher's exact test respectively ($t(198)=3.71$, $p<0.001$, 95% CI=[0.41, 1.36], Cohen's D=0.52; OR=3.18, 95% CI=[1.71, 6.00]).

In terms of the hypotheses predicted for this study, the results confirmed that the unilateral scenario (Figure 14, second panel) was significantly more acceptable when reviewed in isolation than when it was reviewed after the lottery scenario (Figure 14, first panel). A multiple linear regression was undertaken to assess the statistical significance of the

difference with variables representing (1) the difference between the unilateral scenario reviewed in isolation and when reviewed after the lottery scenario (Figure 14, third panel) and (2) the difference between the unilateral scenario reviewed in isolation and the unilateral scenario with submissions (Figure 14, fourth panel). The overall regression was statistically significant ($R2 = 0.04$, $F(2,297) = 5.48$, $p<0.01$). The variable representing the difference between the unilateral scenario reviewed in isolation and when reviewed after the lottery scenario was found to be statistically significant ($\beta = -0.74$, 95% CI = [-1.19, -0.30], $p=0.001$, $\eta2=0.35$). However, while differences between responses to the unilateral scenario with and without submissions were in the direction predicted (in that with submissions, the accused's behaviour was assessed as somewhat less reasonable), the corresponding variable was not statistically significant ($\beta = -0.27$, 95% CI = [-0.71, 0.17], $p=0.23$, $\eta2=0.005$).

*Figure 14. Participants reasonableness ratings of one of the manipulations (first panel) and of the three conditions (second, third, and fourth panels) from Study 11.*




Predictably, an identical pattern was seen with participants' verdicts, with participants much more likely to convict in the unilateral scenario when it was assessed after the lottery

scenario (guilty: not guilty = 90:8) compared to when it was assessed in isolation (guilty: not guilty = 72:30). A logistic regression was undertaken to assess the effect of the two manipulations on verdicts. The effect of seeing the lottery scenario prior to the unilateral scenario was statistically significant (OR=0.21, 95% CI = [0.08, 0.47], $p<0.001$) whereas the manipulation with the prosecution submissions (guilty: not guilty = 75:25) was not (OR=0.80, 95% CI = [0.43, 1.49], p=0.46).

### 5.5.3 Discussion

Study 11 replicated the effect previously seen in Study 9 whereby reasonableness ratings and verdicts in the unilateral scenario became more, rather than less, extreme than the scenario that preceded it. Thus the results were consistent with an asymmetrical order effect whereby a scenario that is assessed as generally objectionable when assessed in isolation becomes even more objectionable when preceded by a more acceptable scenario. This order effect is in the opposite direction to order effects seen in the moral decision-making context whereby a scenario comprising one half of a pairs of similar but opposing scenarios (one acceptable, one objectionable) often becomes less extreme when assessed second. As such, these findings appear inconsistent with consistency type theories that assume that the reason for this effect is that participants wish to appear consistent across the similar dilemmas and therefore alter their responses in the dilemma assessed second to be more akin to the dilemma they have already assessed first.

At the same time, the results do not provide much support for explanations that assume that the order effects are due to the previous dilemmas drawing participants' attention to information that the participants had not previously appreciated as salient. The new condition that included a prosecutor drawing participants' attention to the information assumed to be salient (the fact that a lottery would have been a fairer way to select the individual to be killed) appeared to have very little effect on either assessments of reasonableness or verdicts, though the small effect that there was, was in the direction

predicted. Thus the salience explanation might not be the best explanation, or alternatively the new manipulation might not have been as effective as the previous manipulation. Certainly there are considerable differences between the two. The earlier manipulation consisted of participants actively making a decision on a very gruesome scenario where a very vulnerable crew member lost their life in very extreme circumstances. By contrast, the later manipulation amounted to a relatively brief, bald, and deliberately neutral, statement. This leaves open the possibility that another manipulation that draws the information to the attention of participants in a more effective way might still achieve a similar effect.

Potentially linked to the issue of salience is the issue of the validity of the paradigm, specifically the role of deliberation previously discussed. Commensurate with their seriousness, criminal cases of the type that form the basis of the vignettes in Studies 9 to 11 would invariably be prosecuted before the Crown Court. As such, the decisions would be arrived at by a group of jurors who seek to arrive at a consensus through deliberations that may take days or even weeks to complete. There is good empirical evidence that suggests that groups of decision-makers are more effective than individual decision-makers at identifying salient information (Devine, 2012, p. 180; Ellsworth, 1989, p. 206; Kuhn et al., 1994, p. 295; Mercier, 2010, p. 510; Mercier & Sperber, 2009, pp. 162–163; Sperber & Mercier, 2012, p. 383). Some decisions by groups seem to be more than the sum of the individual contributions. To the extent that the types of order effects that we have disclosed are due to initial failures to take account of salient information, they might not occur when there is deliberation because the group might be more likely to identify the information without the need for a similar preceding decision in which the information is key.

## 5.6 STUDY 12

Though the evidence for salience type theories to explain the phenomenon of order effects in paired legal cases was equivocal, such theories still seemed to be more plausible than other theories previously explored. We therefore looked to take advantage of one of the

characteristic aspects of the experimental paradigm that we had previously not focussed on. As noted above, the seriousness of the types of dilemmas that we were posing to participants would mean that in the real world they would be analysed by a jury. In addition to decisions being made by laypeople, one of the other characteristics of jury decision making is that jurors arrive at a group decision through deliberation. To date, we had analysed decision makers as individuals, in common with much research into jury decision making (Diamond & Rose, 2018, p. 250; E. Greene et al., 2006, p. 240; Hastie et al., 1983, pp. 36, 187; Levett & Devine, 2017, p. 11; N. Pennington & Hastie, 1991, p. 550). However, research suggests that group deliberation is more than simply an averaging mechanism between different views (Devine, 2012, p. 180; Ellsworth, 1989, p. 206; Kuhn et al., 1994, p. 295; Mercier, 2010, p. 510; Mercier & Sperber, 2009, pp. 162–163; Sperber & Mercier, 2012, p. 383). Instead, there is evidence that provided that there is both a diversity of views and the opportunity to debate, then group decision making can be superior to the sum of the equivalent number of individual decision makers (Mercier & Sperber, 2011, p. 63; Sperber & Mercier, 2012, p. 385). There is also evidence consistent with this effect in the context of legal decision-making (Ellsworth, 1989; Kuhn et al., 1994, p. 289; Lagnado, 2021, p. 110; McCoy et al., 1999).

Given our speculation that the asymmetric order effects seen in the pairs of legal cases might be caused by one of the cases highlighting ideas that the participant had not previously realised as salient, we speculated that if this was the explanation, giving participants the opportunity to deliberate as a jury might provide an alternative means to bring about the previously identified effect. Specifically, if the effect was caused by the lottery scenario making participants realise that there was a means of deciding who to sacrifice that would not risk the weakest member of the group being killed at the expense of the strongest (ie, a lottery), deliberation as a group might be more likely to reveal this idea to jurors than if they considered the dilemma independently. We therefore sought to see if we could replicate the effect shown in Studies 9 and 11 using a condition whereby the jurors had an opportunity to deliberate as part of a group before assessing the unilateral scenario and a condition whereby individual jurors assessed the same scenario without such an opportunity.

Our prediction was that if the asymmetrical order effect previously identified was caused by participants realising that there was a preferable means of selecting the sailor to be sacrificed than that actually chosen by the sailors and casting their behaviour in a more negative light, then participant jurors who had had an opportunity to deliberate in groups would be more likely to have settled on this information, and correspondingly their responses would be more in line with the participants who had previously assessed the lottery scenario in Studies 9 and 11.

## 5.6.1 Method

### 5.6.1.1 Participants

One hundred and fourteen participants were recruited from students studying an undergraduate psychology methodology course at University College London. Of these, 85% were female and 15% were male. Ages were from 18 to 23 (M=18.7, SD=0.86). The study was calculated to have an 80% power to detect an effect size of d=0.61. Participants participated as part of the methodology course and wrote up the experiment and results, but were not informed of the theoretical background or predictions prior to participation. Participants were not financially remunerated for their participation. The study was conducted in accordance with approval obtained from UCL's Research Ethics Committee (EP/2018/005). Informed consent was obtained from each individual in advance of participation by providing them with study information sheet and a consent form to agree to.

### 5.6.1.2 Design

We used a between participant design with one independent variable with two levels in which all participants assessed the unilateral scenario previously used. Participants were

randomly assigned to 2 conditions: an individual condition where participants assessed the scenario in isolation without the ability to discuss with others; and a group condition where participants assessed the scenario after having an opportunity to discuss the scenario with other members of a group. Whereas juries in England and Wales generally amount to 12 individuals, given the relatively small pool of available participants, it was decided to compromise at groups comprised of 6 individuals. In order to ensure a realistic prospect of an effect due to group membership, a greater overall proportion of students were assigned to the group condition so that a sufficiently large number of groups could be formed. A final ratio of 2:1 of individuals : groups (or 1:3 of participants as individuals: participants as group members) was chosen as a compromise designed to ensure both appropriate representation of participants as individuals and as group members, as well as a sufficiently large number of overall groups.

#### 5.6.1.3 Materials

Participants in all conditions were given written instructions advising them that they would be asked to put themselves in the place of a juror in a Crown Court to decide a hypothetical case. It was explained to them that in law, an accused who would otherwise be convicted of an offence can avoid liability if they have a defence. The defence of duress of circumstances was described as where an accused commits what would otherwise be a life in order to save a life or prevent serious harm to somebody. Participants were told that they would be presented with a case where the accused argue duress of circumstances and asked whether they thought the accused should be convicted or acquitted.

Participants were given a written document summarising the relevant facts of the scenario as described in Studies 2 and 4. In summary, this was that a small 3-man crew had been shipwrecked on a lifeboat without communication equipment and had exhausted their rations over a period of 2 weeks. The captain and mate secretly discussed the situation and decided that they might survive for a little longer if they resorted to cannibalism. They agreed

to kill the apprentice who they felt would die soonest. After he fell asleep, the mate held him down while the captain killed him. They subsequently consumed him. Participants were told that the two were being prosecuted for murder, but relying on the defence of duress of circumstances. This was described as where the accused committed the offence because they reasonably believed that they would die or be seriously injured and a person of reasonable firmness would have acted in the same way.

All participants were asked individually to give a verdict of guilty or not guilty, to choose a reasonableness point on a 7-point Likert scale, and to give reasons for their decision. Participants in the group condition were additionally asked to complete the same measures by consensus after deliberating but prior to responding individually.

#### 5.6.1.4 Measures

All participants completed a single form which asked them to indicate a verdict of guilty or not guilty; reasonableness on the same 7-point Likert scale previously used (1 = completely unreasonable / 2 = unreasonable / 3 = somewhat unreasonable / 4 = evenly balanced / 5 = somewhat reasonable / 6 = reasonable / 7 = completely reasonable); and to explain the reasons for their decision. Participants were also asked demographic details comprising their age and gender.

Participants in the group condition were additionally asked to complete the same form as a group by consensus after deliberating but prior to completing the same measures as an individual. However, this group form did not request demographic information.

#### 5.6.1.5 Procedure

Participants were randomly assigned to either the individual or the group condition. Participants assigned to the group condition were randomly assigned further to a small group of 6. Those in the individual condition completed the study in an undergraduate laboratory in silence, supervised by university demonstrators. Those in the group condition participated together with their assigned group in private rooms. The demonstrators administered the survey, but were excluded from the private rooms while the participants deliberated and completed the measures.

In the individual condition, participants were provided with an envelope containing the instructions, participant information sheet, consent form, materials, and measures. Participants were instructed to read the instructions, the participant information sheet, and complete the consent form if they were happy to participate. They were then asked to read the facts of the case and complete the form giving their verdict, reasons, and demographic information. Participants were instructed to return all materials other than the participant information sheet. Demonstrators ensured that all materials were collected from the participants before they left.

Those participants in the group condition were additionally instructed to discuss the case, agree on a group verdict, and to provide reasons on behalf of the group before they completed the individual measures described above. Demonstrators similarly ensured that all materials were collected from the participants in the group condition before they left, with the exception of the participant information sheet.

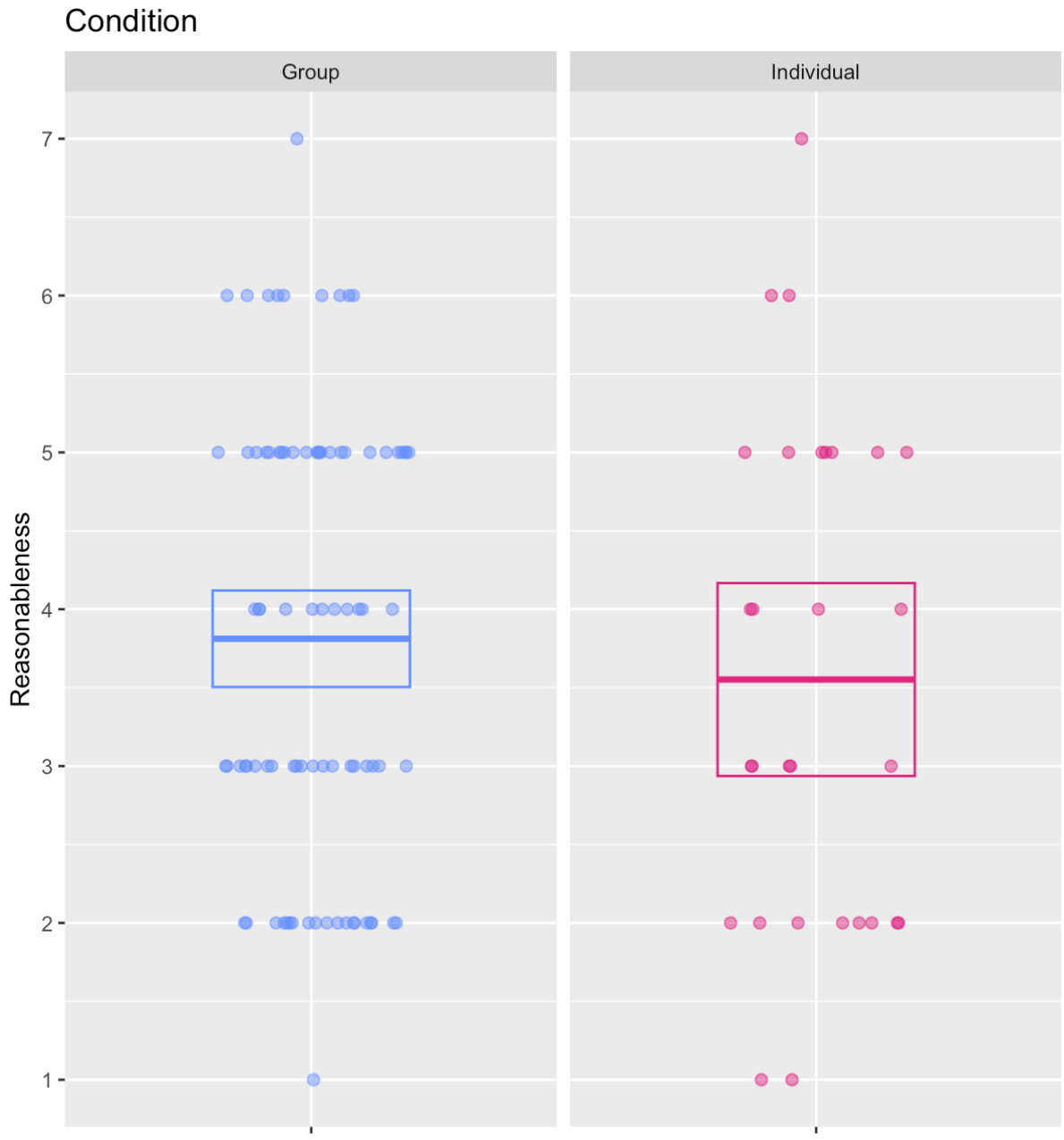
All participants participated in the study on the same afternoon. All individual participants began the study at the same time. Participants in the group condition were given staggered half-hour periods to attend the private rooms to participate.

### 5.6.2 Results

Of the one hundred and fifteen people participated in the study. One individual was excluded on the basis that they were familiar with the case of *R v Dudley v Stephens*. 29 people participated as individuals and 85 participated as members of a group. Those in the group condition were assigned to 15 groups of 6, but due to absences, some groups had fewer than 6 members. On the day, 11 groups comprised 6 members with 3 groups of 5 members and 1 group of 4 members.

In terms of the individual verdicts given by jurors who considered in isolation compared to those who deliberated as a member of a group, the proportions were uncannily similar. Those in the individual condition delivered verdicts in a ratio of 21:8 guilty:not guilty, a ratio of 0.72 when rounded, whereas those in the group condition delivered verdicts in a ratio of 61:24 guilty:not guilty, also a ratio of 0.72 when rounded. Obviously, the tiny difference between the two groups was not significant. A Fisher's exact test confirmed this ($p=1$, $OR=0.97$, 95% $CI=[0.33, 2.68]$). There was somewhat more of a difference between individual reasonableness assessments in the conditions, but this was not in the direction predicted, with individual reasonableness assessments in the individual condition rating the behaviour as slightly less reasonable ($M=3.55/7$, $SD=1.62$) than in the group condition ($M=3.81/7$, $SD=1.43$). A Welch two-sample t-test indicated that this difference did not reach statistical significance ($t(43.8)=0.77$, $p=0.45$, Cohen's $D=0.18$).

*Figure 15. Reasonableness ratings by jurors deliberating prior to decision as a group compared to those deciding as individuals in Study 12.*




An interesting incidental finding of the study was that there was a strong correlation between verdicts given by the group and individual verdicts given by members of that group.

That is, where a group arrived at a particular group verdict, almost all of the individual members were also likely to give that verdict when asked individually. Thus, of the 64 individuals who were in groups that collectively arrived at a verdict of guilty, only 7 subsequently gave an individual verdict of guilty; and of the 21 individuals who were in groups that collectively arrived at a verdict of not guilty, only 4 subsequently gave a verdict of guilty. Unsurprisingly, a Fisher's exact test confirmed a very strong correlation between group verdict and individual verdict ($p<0.001$, OR=32.1, 95% CI = [7.8, 171.7]). While some correlation would have been expected between individual verdict and group verdict (because the majority in a group would be assumed to have a greater influence on the final verdict), the correlation seemed to be much higher than would be expected. This suggested that either (1) individuals who would otherwise have preferred a different verdict had been persuaded to take a different position by the group discussions; and/or (2) that individuals who had agreed on a unanimous group verdict subsequently gave the same individual verdict to appear consistent, even if they would otherwise have preferred a different verdict.

### 5.6.3 Discussion

The headline statistical findings of this study suggest little effect of being in a group deliberation on participants' decisions, but a deeper examination of the decision-making patterns suggests that there may also be more complicated factors involved. Considering the hypothesis that those in the group condition should be more likely to realise that there was a fairer way to select those to be killed and would therefore be more likely to convict, the statistical analysis did not bear this out. The conviction rate for those in the individual condition was very similar to the rate for those in the group condition. Similarly, the reasonableness ratings for participants who considered the scenario was very similar for those in both conditions, with those in the group condition assessing the behaviour of the putative accused as slightly more reasonable. However, it might be prudent to take a little care with these headline findings given there was evidence that other factors might have been in play. In particular, the distribution of individual verdicts within the groups showed a clear bimodal

distribution with individual verdicts very strongly influenced by the eventual agreed verdict of the group. Individuals were often unanimously in agreement as individuals with the group verdict. Given that participants were randomly assigned to a particular group, there ought not to have been a bi-modal distribution unless there were further processes in play within the group.

There seemed to be at least two possible additional processes at play. One was a consistency effect in line with the research in Section 4. It was conceivable that once participants had agreed on a majority verdict, when they were asked for their individual verdict, they could have been at least partially motivated to give a verdict consistent with the majority verdict rather than the verdict that they would have preferred had they not been required to reach a group verdict by majority.

A second possibility was that there were other factors than the method of selection that the group identified as salient and which subsequently persuaded the group one way or another. From previous research we know that jury, or group, decision making is more than the sum of its parts (Devine, 2012, p. 180; Ellsworth, 1989, p. 206; Kuhn et al., 1994, p. 295; Mercier, 2010, p. 510; Mercier & Sperber, 2009, pp. 162–163; Sperber & Mercier, 2012, p. 383). We also know that the majority in jury decision making does not always prevail. In some cases, an argument from the minority succeeds in persuading the majority. Furthermore, while we had predicted something akin to a 'eureka' moment whereby participants realised that there was a particular piece of information (a fairer way of selecting the victim that would be less likely to result in the death of the most vulnerable), the present scenario presented a fairly complicated set of facts. Therefore, unlike other research where there is only one solution (Duncker, 1945), the complex background matrix of facts might have given rise to a number of different pieces of information that may have been treated as relevant by the groups. Thus, the group verdict could have been influenced by more than one piece of information, and that information could have swayed the group in both directions.

The lack of a recording or transcript of the group deliberations compounded the

difficulties in diagnosing what additional factors might have influenced the group deliberations. Had this information been available, a qualitative assessment of the information that influenced the debate may well have shed some light on the debates. An additional issue was the relatively small jury size necessitated by the limited sample size. Though most jury groups in the study consisted of 6 members, this is still half the size of the typical jury. It is difficult to know the extent to which this influenced the nature of the debates and whether there is a minimum group size to achieve the discovery or 'assembly bonus' effects seen in group deliberation (Mercier & Sperber, 2011, p. 63). Other factors relevant to the design include the specified time slots given to the juries which may have exerted some pressure to arrive at a premature consensus. Real-life juries, by contrast, are not given a deadline and deliberations may continue for a considerable time.

The unexpected influence of the groups is an interesting phenomenon worthy of further exploration, but it complicates the search for an explanation of the order effects seen in these paired dilemmas. Given the uncertain effect of the use of group deliberation as a manipulation, it is difficult to draw firm conclusions regarding the competing theses we are examining.

## 5.7 STUDY 13

For Study 13, we returned to the individual decision-making paradigm previously used and sought to address the potential issue with Study 11 of the manipulation being too weak with the use of a more colourful manipulation. In Study 11, we replicated the order effect previously evidenced in Studies 9 and 10 whereby participants who had previously determined a decision in which the sailors adopted a fairer selection system that did not lead to the death of the weakest member looked less favourably upon the behaviour of sailors who had unilaterally selected the weakest member of the group for death. However, the attempt to elicit the same effect using different means in Study 11 had proved to be unsuccessful. The additional condition deployed to achieve this in Study 11 was to have the prosecutor raise the

idea in the closing submissions. In that condition, the information was presented in a neutral and descriptive manner. This raised the possibility that the manipulation was not strong enough compared to the process of actually undertaking a decision on a lively and thought provoking dilemma.

To address this potential issue, In Study 13 the condition whereby the prosecutor disclosed the information in closing submissions was repeated, but with two key changes. First, the example chosen for the prosecutor's submissions to convey the information was an intriguing example taken from sailing history. This was the history of the wreck of the American whaling ship Essex. In 1820, the ship was sunk by a Sperm Whale while 2,500 miles off the coast of South America. After the ship was destroyed, the crew put to sea in small whaleboats which then became separated. The captain, George Pollard's whaleboat survived for 2 months at sea before the crew ran out of food and began to die. Initially the surviving seamen cannibalised the bodies. Once the bodies of the seamen who had died of natural causes had been consumed, the crew decided to draw lots to decide who would be killed for the survival of the others. Owen Coffin, the captain's 17-year-old cousin drew the black spot. The captain had sworn to protect Coffin and offered to protect him, but Coffin reportedly said 'No, I like my lot as well as any other.' Further lots were then drawn to decide who would be the one to kill Coffin. Coffin was killed and the others consumed his body. Two sailors eventually survived. The second change was that participants in the survey would be explicitly asked after the survey what they might have done differently to assess the extent to which they had synthesised this information, and also to assess the extent to which there was a difference in understanding between the conditions.

If salience theories best explain the order effects that we have seen, our prediction would be that participants in the condition where the prosecutor highlights the option of a fairer means of selecting the sailor to be killed should assess the behaviour of the sailors as less reasonable and correspondingly be more likely to convict.

## 5.7.1 Method

### 5.7.1.1 Participants

Two hundred participants were recruited using the online survey platform Prolific on the basis of British nationality and residency in England and Wales and successfully completed the survey. Of these, 130 (65%) were female and 70 (35%) were male; ages were from 18 to 71, M=37.3, SD=12.1; 15% were students; and 53% were in full-time employment, 17% in part-time employment, and 30% were unemployed or of other status. The study was calculated to have an 80% power to detect an effect size of d=0.40. Participants were paid £0.60 for their time. The study was conducted in accordance with approval obtained from UCL's Research Ethics Committee (EP/2018/005). Informed consent was obtained from each individual in advance of participation by providing them with study information and a consent form to agree to.

### 5.7.1.2 Design

We used a between participant design with one independent variable with two levels in which all participants viewed the target unilateral scenario. Participants were randomly assigned to 2 conditions: a control condition where after reviewing the scenario, participants read neutral submissions by the prosecution that essentially described the facts and the law; and an experimental condition where after reviewing the scenario, participants read much more colourful submissions by the prosecution, referring to the facts of the shipwreck of the Essex and how the survivors had fairly selected who to kill using a lottery.

### 5.7.1.3 Materials

Participants in both conditions were given written instructions advising them that they would be asked to put themselves in the place of a juror in a Crown Court to decide a hypothetical case. It was explained to them that in law, an accused would otherwise be convicted of an offence can avoid liability if they have a defence. The defence of duress of circumstances was described as where an accused commits what would otherwise be a life in order to save a life or prevent serious harm to somebody. Participants were told that they would be presented with a case where the accused argue duress of circumstances and asked whether they thought the accused should be convicted or acquitted.

Participants read a document summarising the relevant facts of the unilateral scenario as described in Studies 2, 4, and 5. In summary, this was that a small 3-man crew had been shipwrecked on a lifeboat without communication equipment and had exhausted their rations over a period of 2 weeks. The captain and mate secretly discussed the situation and decided that they might survive for a little longer if they resorted to cannibalism. They agreed to kill the apprentice who they felt would die soonest. After he fell asleep, the mate held him down while the captain killed him. They subsequently consumed him. Participants were told that the two were being prosecuted for murder, but relying on the defence of duress of circumstances. This was described as where the accused committed the offence because they reasonably believed that they would die or be seriously injured and a person of reasonable firmness would have acted in the same way.

In the control condition, the prosecution closing speech was neutral and descriptive, summarising the facts and the law. In this speech, the prosecution simply asserted that the actions of the accused were not reasonable and therefore they should be convicted. In the experimental condition, the prosecution closing speech also asserted that the actions of the accused were not reasonable, but supported this assertion by reference to the facts of the case of the Essex. The submissions explained that in that case, the crew had decided to draw lots, and when a junior member of the crew, Coffin, was selected, the captain had offered to protect him, but Coffin had refused to allow him. Participants were told that the crew then drew lots again to decide who would execute him.

All participants then read the same defence submissions and the judge's summary of the law where the law including the burden of proof was explained. Participants were advised in the summary of the law that the only issue was whether what the accused did was reasonable and if they thought that the accused's action were, or might have been, reasonable, they should find them not guilty.

All participants were asked individually to give a verdict of guilty or not guilty, to choose a reasonableness point on a 7-point Likert scale, and to give reasons for their decision as before. Additionally in this study, participants were subsequently asked to explain what (if anything) the accused have done differently when faced with these circumstances. Participants were asked to identify the issues in the case, to state how easy they found the explanation of the law, and whether they had any relevant legal experience or knowledge.

#### 5.7.1.4 Measures

All participants responded to the same measures. These were a binary indication of verdict which was either guilty or not guilty; a gradated Likert response with 0.1 increments from 1-7 where each number was matched with a verbal description (1 = completely unreasonable / 2 = unreasonable / 3 = somewhat unreasonable / 4 = evenly balanced / 5 = somewhat reasonable / 6 = reasonable / 7 = completely reasonable); and a text response box to give an explanation of their decision.

After completing their decision, they were asked what the accused might have done differently (if anything) when faced with this situation and provided with an open text response box to explain.

Once participants had responded to one scenario or both, all participants responded to a series of other measures. An attention check invited the participants to identify the issues in

the case with 2 correct answers (whether the accused should be acquitted or found guilty, how reasonable it was for the accused to kill when faced with starvation) and one incorrect answer (whether the accused's lives were at risk). Participants were also asked if they found the explanation of the law easy or difficult, and if they had any relevant legal knowledge or experience.

#### 5.7.1.5 Procedure

As with Study 4, participants were recruited online from the Prolific platform and participated in the survey using the online platform Qualtrics in a place of their choosing, using their own device. On initial referral, participants were first given with the study information form. They were then asked to complete the consent form comprising a number of statements to which an affirmative answer was required in order to participate in the survey. An anonymous user identification was collected to enable subsequent matching of demographic data without compromising the participants' anonymity.

Participants were randomly assigned to one of the two conditions by the Qualtrics platform in a balanced way so that half were in the control condition with the neutral prosecution submissions and half were in the experimental condition with the prosecution submissions explaining about the shipwreck of the Essex and the procedure adopted by the shipwrecked sailors in that event. After viewing the scenario, and the prosecution and defence submissions, participants were first asked to give a verdict, secondly to assess the reasonableness of the accused's actions, and thirdly to give reasons for their decision.

Once participants had completed their responses to the scenario, they were then asked what the accused might have done differently and then the additional measures.

After completing the survey, participants were thanked for their participation and referred back to the Prolific survey platform to confirm their participation. Once both

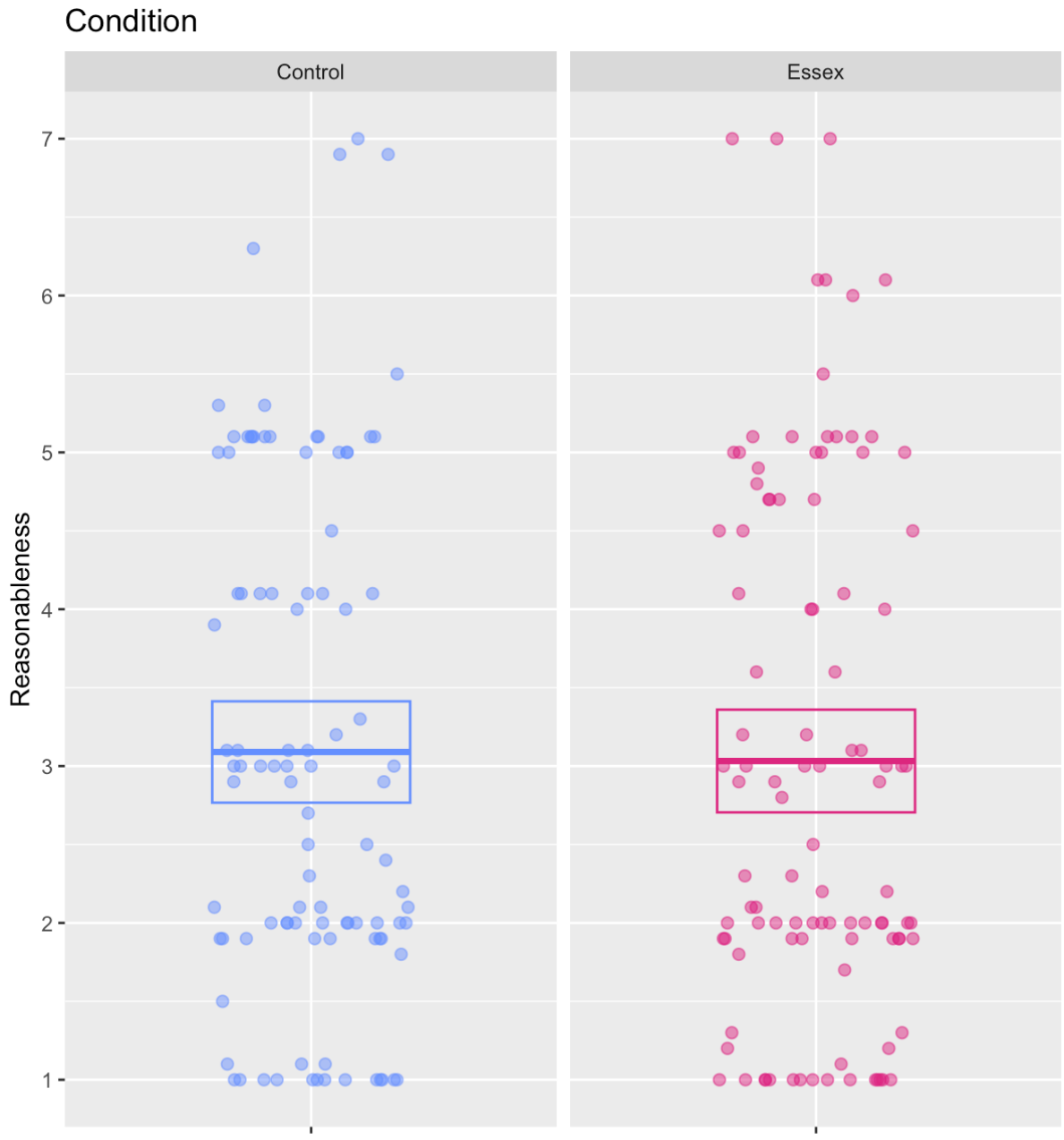
platforms had confirmed the participant's successful completion of the survey, their remuneration was authorised.

### 5.7.2 Results

200 participants successfully completed the survey. Of these, 84% found the law easy to understand with only 16% finding it difficult. In terms of issues identified, only 1% of participants exclusively selected the legally incorrect issue of whether there was a risk to the life of the accused, and 8% selected this in combination with one or more of the other issues. This suggested a generally good understanding of the law as though the issue of whether there was a risk to the lives of the accused was a condition for the availability of the defence of duress of circumstances, it was not an issue on the facts presented as the prosecution had accepted that there was a risk to the accused's lives. 7% of the participants had some legal training or experience, but this was universally limited and did not suggest any familiarity with the precedents referred to in the materials. As such, no exclusions were made.

In terms of the planned contrasts, contrary to the predictions of salience type theories, there was little difference between the conditions, see Figure 16. Those in the experimental, Essex, condition were slightly more likely to convict (68%) than in the control condition (63%). While in the predicted direction, this small difference did not reach statistical significance according to a Fisher's exact test (OR=0.79, 95% CI=[0.42, 1.47]). Likewise, participants assessed the accused's behaviour in the experimental Essex condition as very marginally less reasonable (3.03/7) than in the control condition (3.09/7), but this modest difference did not approach statistical significance on a t-test (t(198)=0.25, p=0.80, 95% CI=[-0.40, 0.52], Cohen's D= 0.04).

*Figure 16. Individual reasonableness assessments by control condition given no information and condition advised of the facts of the shipwreck of the ship Essex in Study 13.*




Notwithstanding the lack of an apparent difference between the conditions, it seemed

that Participants in the Essex experimental condition were much more likely to be aware of the idea of adopting a fairer means of choosing the victim compared to in the control condition. The open text responses to the question about what the accused might have done differently were coded by a coder blind to the experimental conditions. Those participants who either referred to adopting a form of lottery, or to discussing the situation to reach a consensus were coded as having identified the possibility of a fairer selection process. Other responses that were coded as not having identified this possibility included responses such as that the accused should have waited longer, sought other means of sustenance, or simply refrained from cannibalism. Only 9% of those in the control condition were coded as having identified the possibility of a fairer means of selection compared to 37% of the experimental condition. This difference was statistically significant pursuant to a Fisher's exact test (OR = 5.82, 95% CI =[2.54, 14.68], $p<0.001$), indicating the manipulation in the experimental condition had successfully communicated the information to a significantly higher proportion of participants in that condition.

### 5.7.3 Discussion

Study 13 addressed the concern from Study 11 that the alternative manipulation explicitly disclosing the information regarding a fairer means of selection of the victim in the shipwreck cases was not as strong as that in Studies 9, 10, and 11 where participants previously made a decision in a case where a fairer means of selection was adopted. The analysis indicated that a much larger proportion of participants were aware of the idea in the experimental condition where the prosecutor had disclosed that there was a fairer means of selection than in the control condition where the prosecutor did not suggest the idea. Nonetheless, this information appeared to make little or no difference to participants' responses between the conditions. Salience type theories would predict that if the order effects seen in previous experiments were due to previous scenarios drawing the attention of participants to the possibility of a fairer means of selection, which thus cast a more negative light on the behaviour of the accused, the disclosure of the information should have had a

similar effect on participants' responses, such that the experimental condition should have seen more convictions and lower reasonableness ratings. This is not what was seen. Studies 11 and 13 thus seem to be inconsistent with this theoretical explanation.

Though it remains a possibility that the new manipulation used in Studies 11 and 13 was not sufficiently strong compared to the original manipulation used in Studies 9, 10, and 11, this seems unlikely. Though we did not test it explicitly in the original manipulation, the fact that participants had to actively make a decision on the basis of the information disclosing a fairer means of selection of the victim, one might assume that a very high percentage of participants were aware of this counter factual when they came to decide on the next dilemma. In the new condition, only around a tenth of participants thought of the possibility of a fairer means of selection compared to slightly more than a third in the experimental condition. As such, even a small effect size ought to have been identified given the sample size. There is also the possibility that the lack of an effect might have been a Type II error, but again this seems improbable.

Overall, Study 13 appears not to provide support for the theory that was most compatible with our experimental findings prior to this study. As noted above, the asymmetrical nature of the order effects in Studies 8, 9, 10, and 11 and the fact that the direction of effect where responses less became more dissimilar to previous responses in Studies 9, 10, and 11 seemed *prima facie* incompatible with consistency type theories. Up until Study 13, it seemed that salience type theories might be more compatible with the experimental findings on the assumption that the manipulation in Study 11 was insufficiently strong to show an effect. However, in the light of the Study 13, this theory also seems incompatible with the experimental findings.

## 5.8 GENERAL DISCUSSION

The goal of this section was to seek to replicate the order effects seen in moral

decision-making in the context of legal decision-making, and the studies undertaken demonstrated emphatically that the same phenomenon exists. The project also went beyond the order effects seen in the moral decision-making context where the direction of the order effects in that research field tend to be attractive, such that responses to a scenario presented second become more similar to the scenario presented first. By contrast, we found that there were some pairs of scenarios where there was an apparent repellent effect whereby responses to the scenario presented second became more dissimilar to responses to the first scenario. Whether the effect was attractive or repellent, we also saw clear evidence of the types of asymmetric effects found in previous moral decision-making research whereby responses to one of the two scenarios remained relatively stabile regardless of the order in which it was presented and responses to the other scenario were relatively labile, changing substantially depending on whether the scenario was presented first or last. Thus in each of the studies where we sought to demonstrate or replicate an order effect (Studies 8, 9, and 11) we obtained significant results. The exception was Study 10 where the effects were in the direction predicted, but did not reach statistical significance. Study 8 indicated an orthodox attractive order effect between the scenarios, whereas Studies 9, 10, and 11 demonstrated a novel repellent order effect. All of the order effects seen were asymmetrical.

While the phenomenon has been fairly robustly demonstrated empirically, a theoretical explanation remains elusive. Perhaps the most popular explanation, that participants strive to maintain an appearance of consistency between the paired scenarios, is fairly inconsistent with our findings, at least as the primary explanation. The asymmetrical nature of the findings, both in previous research, and our findings, requires additional changes to the auxiliary hypotheses, for example that it is more difficult for participants faced with some scenarios to change their responses to the second scenario to approximate their responses to the first scenario without appearing unreasonable. The moral decision-making example of Transplant might illustrate the type of scenario, as no reasonable person would seriously suggesting murdering a patient, even if it would save several lives overall. However, the new effect that we have identified whereby responses in the second scenario become less akin to responses in the first scenario piles further pressure on the theory,

requiring either further changes to the auxiliary hypotheses, or abandoning consistency theories as the most plausible explanation of these patterns. Salience-type theories that suggest that the effects are the result of some scenarios drawing the attention of participants to particular factors that they had overlooked fare slightly better, but are still not wholly consistent with our findings. Salience theories are consistent with Studies 8, 9, and 10 and the replication condition in Study 11 in that it is plausible that some piece of information highlighted in the scenario presented first (such as the risk that a vulnerable individual will be taken advantage of or the possibility of a fairer means of selection) caused participants to determine the second dilemmas differently compared to had they been presented in isolation. However, the new condition in Study 11 and Studies 12 and 13 do not appear terribly consistent with salience explanations. If salience explanations are correct, it would be reasonable to assume that it would also be possible to trigger a similar pattern by explicitly communicating the salient information to the participants. Studies 11, 12, and 13 appear to belie this though: disclosing the relevant information in a neutral way (Study 11); in a more colourful way (Study 13); or facilitating the discovery of this information by allowing group deliberation (Study 12) appeared to have negligible effect on responses. Similarly to coherence explanations, it is possible to posit other auxiliary hypotheses, such as stochastic factors, the greater effectiveness of actively using information to take a decision compared to passively receiving the information, or other countervailing effects, but at present these are fairly speculative explanations. Given that further empirical work is inevitable, it would be invaluable to have some theory to guide this empiricism, so further work is needed here.

Turning to empiricism, it should be noted at the outset that the various paradigms used in trolley-type moral psychology research seem inherently problematic. Trolley problems with their binary outcomes and simplified dilemmas designed to test philosophical theories such as deontology ('though shalt not kill') against utilitarianism ('maximise utility') were assumed to be sufficiently simple to test theories in limited environments, but human rationality increasingly seems significantly more sophisticated than assumed. While these scenarios generally have a binary outcome for participants to determine, there appear often to be more than two cognitive processes taking place. Furthermore, there remains something of

an issue with the design often used in moral decision-making trolley experiments and in some of the studies that have formed part of our research. The 'design' originally adopted in the moral decision-making context and subsequently adopted by us to attempt to replicate the effects seen was not initially used to test a particular theory, but rather a standard approach of randomised order to minimise the effect of unknown stochastic primacy or recency effects, effects that were not originally of experimental interest. As we know from the asymmetrical order effects identified, this approach did not eliminate all additional effects beyond those being tested in the experiments. At the same time, while the effects seen are very interesting, this design is not ideal for isolating individual effects because randomising the order gives rise to more than two different conditions. For example, with two dilemmas A and B presented in a random order, there are a number of different conditions: A in isolation; B in isolation: A after B; and B after A. The result is that this limited design makes it less straightforward to identify the causes of the identified effects.

Given the limited state of the theory, the field of future experiments that would be advisable seems fairly open. The replication of the phenomenon identified with real-life lay or professional adjudicators is always valuable, particularly as lay participants are not likely to be familiar with the concept of *stare decisis*. However, given the cost and difficulty of such research and the limited insight that we currently have, the priority might be to gain a better understanding of the phenomenon we have identified with lay participants first. To date, we have conducted a series of experiments based primarily on two different scenarios, one civil and one criminal, but a greater diversity of scenarios might provide greater illumination. Furthermore, given the nature of legal proceedings where a variety of different adjudicators contribute to the development of the law, the most interesting findings would be if institutional order effects could be demonstrated across different decision-makers rather than individual order effects by individual decision-makers as we have shown here. For the reasons explained in the previous paragraph, the standard paired dilemma paradigm should probably be replaced by more rigorous experimental designs.

The phenomenon identified is potentially relevant to real-life legal decision-making,

but in a slightly different way to moral decision-making. In the moral decision-making field, order effects are mainly seen as concerning due to the implication that moral decision-making can be influenced by arbitrary effects. It is generally assumed that moral decision-making should be guided by a stable set of underlying values and principles that should be immune from irrelevant influences such as order (A. B. Moore et al., 2008, p. 556; Schwitzgebel & Cushman, 2012, p. 136). In the legal context, there is more leeway granted to decision-makers, in part due to a recognition of how difficult it can be to establish a legal principle that generalises to all cases likely to come before the courts (Beale, 1916, p. 147; Fuller, 1957, pp. 667–668; Goff, 1999, p. 318; Hart, 1961, p. 128; Wambaugh, 1894, p. 56). Nonetheless, the possibility that legal decision making might be influenced by apparently arbitrary factors such as the order that cases come before the courts is also problematic. At present, this phenomenon is not yet well understood or theoretically explained. This makes it difficult to draw definite policy prescriptions. The paradigms used in our studies focussed on individual inconsistency, but in real-life it is rare that a single decision-maker establishes all the precedents. Rather, it is more common for different decision-makers, or at least different panels of decision-makers to establish the precedents, making the question of institutional inconsistency also relevant. Much also depends on the theoretical explanation for the phenomenon as some explanations would be problematic and others unproblematic. For instance, it would be problematic if the effects seen were the product of cognitive shortcomings whereas it might be less problematic if the effects were explained by some sort of salience theory whereby some cases made the decision-makers take into account factors that they had not previously recognised as relevant. More modestly, the order effects demonstrated raise some cautions about the sufficiency of simply randomising the order of scenarios to address any potential collateral primacy or recency effects: the order effects seen, including the asymmetrical nature of the effects, suggests that there may often be more complex factors at play.

# 6. THE EFFECT OF CAUSAL INFORMATION ON OUTCOMES

## 6.1 INTRODUCTION

A third category of circumstances where legal decision making may be influenced by factors that are not well explained by existing psychological theory is where decision makers are provided with explanations that purport to account for offending behaviour. Decision-makers in criminal courts generally apply a common-sense view of human psychology where the accused is treated as responsible for their own, self-originating, choices (R. J. Allen, 2000; Hart & Honoré, 1985; Morse, 2004). Yet this common-sense view is being increasingly disrupted by scientific research that is identifying factors that appear to influence propensity to offend. If offending behaviour can be explained by identifiable causes, many feel that it is less reasonable to hold the accused fully responsible for that behaviour (Ayer, 1946, pp. 276–277; Dennett, 1984, p. 157; G. E. Moore, 1912, p. 111; Smart, 1961, p. 293).

One of the most plausible candidates for an identifiable marker of propensity to offend is the Monoamine Oxidise A ('MAOA') genotype × environment interaction. The interaction is a product of (1) whether an individual has a genotype that codes for low levels of the production of MAOA, an enzyme that breaks down neurotransmitters, and (2) whether the individual suffered childhood abuse. Research suggests that men with lower levels of MAOA who suffered childhood abuse are much more likely to react to perceived provocation and to be convicted of a violent offence (Byrd & Manuck, 2014; Caspi et al., 2002; Kim-Cohen et al., 2006). Caspi et al. (2002), for example, assessed MAOA activity in 499 males (at age 26) from the Dunedin longitudinal study (96% of the living cohort members). They found that individuals with lower levels of MAOA activity who had been maltreated made up only 12% of the sample, but they were responsible for 44% of the violent convictions recorded in that sample.

The plausibility of the research into MAOA has meant that it has often met the threshold for admissibility of scientific evidence in criminal proceedings, and the evidence

indicates that it is being increasingly adduced in many jurisdictions (Catley & Claydon, 2015; de Kogel & Westgeest, 2015; Farahany, 2015; Mcswiggan et al., 2017). To date, an in accordance with widespread intuition, the research has generally been taken to reduce the responsibility of wrongdoers and has correspondingly been put forward by defence lawyers in mitigation of sentence (Catley & Claydon, 2015; de Kogel & Westgeest, 2015; Denno, 2015). Anecdotal evidence suggests that this practice has met with some success (Feresin, 2009; Feresin & Owens, 2011). Given this real-world background, it might be assumed that experimentally isolating responsibility reducing effects of causal information might be relatively straightforward, but this has not been the case. While there has been some evidence suggesting effects on reducing offence and sentence seriousness, this has been far from unequivocal.

Less attention has been paid to verdicts than sentence, but some experimental research suggests that causal information reduces conviction rates. Confer and Chopik (2019) found that putting offending down to a brain tumours and childhood abuse reduced conviction rates compared to controls. Gurley and Marcus (2008) similarly found that participants were more likely to choose not guilty by reason of insanity than guilty when defendants were described as having a psychotic disorder or brain injuries. By contrast, Berryessa et al (2021) found little effect of different types of neurobiological evidence including genetics on sentence. Similarly, Schweitzer & Saks found little effect of neuroimaging on verdict (2011), though neuroimaging alone has rarely been found to influence outcomes (Aono et al., 2019, p. 16; D. A. Baker et al., 2013; LaDuke et al., 2018; Mowle et al., 2016, p. 737).

Consistent with real world practice, more empirical attention has been paid to effects on sentence than verdict, and this has resulted in some supporting evidence. Aspinwall et al found that professional judges imposed a slightly shorter sentence when a putative accused's behaviour was explained by MAOA genotype and childhood abuse (Aspinwall et al., 2012), but subsequent efforts to replicate this effect in Germany and with lay participants based on the Aspinwall materials have not proven significant (Fuss et al., 2015; Guillen Gonzalez et al., 2019; Remmel et al., 2019). A small survey about the effect of autism and genetics

suggested that most professional judges thought this information was mitigating (Berryessa, 2016). Greene and Cahill found that neurobiological evidence reduced the proportion of defendants recommended for a death sentence (E. Greene & Cahill, 2012). Notably, it was high risk defendants who were less likely to receive a death sentence when this evidence was presented. The research of Saks et al (2014) similarly suggested that such evidence reduced death sentences. Gordon and Greene (2018) found that information about MAOA genotype + mistreatment led to fewer death sentences, but information about genotype alone actually increased death sentences. Kopel et al also found that doubt in free will reduces support for retributive punishment (Koppel et al., 2018). Muir (2019) looked at MAOA genotype, suggesting that participants viewed defendants as both more dangerous and less culpable, imposing shorter sentences overall. Kim et al (2015) found an interesting pattern in that where the accused was from an abusive family, MAOA genotype information reduced sentences, but when the accused was from a loving family, MAOA genotype information actually increased sentences.

However, other research has found little effect of such causal information, or has found aggravating effects. Research by Appelbaum & Scurich suggested that genetic and genetic + abuse explanations were aggravating (Appelbaum & Scurich, 2014). Likewise, Robbins & Litton (2018) found little difference other than an apparent aggravating effect of a genetic condition. Similarly, Appelbaum et al (2015) saw no effect of genetic or neuroimaging on serious trial outcomes and Scurich & Appelbaum (2015) found no effect of genetic evidence on less serious offences either. Studies by Lynch et al (2019) indicated that genetic and environmental background did not influence punishment decisions, but it did influence evaluations of free will. Relatedly, Lui et al (2019, p. 479) found that genetic explanations mitigated culpability but nonetheless aggravated sentencing severity for defendants identified as psychopaths. Other research has seemingly found little effect of neuroscience (Blakey & Kremsmayer, 2018; Marshall et al., 2017; Mowle et al., 2016, p. 737).

This indeterminate evidence has led many to conclude that there is no effect of

MAOA genotype or environment on either verdict or sentence (Aono et al., 2019; Scurich & Appelbaum, 2015, 2017). However, others have suggested that the mixed experimental results may be because the information is a 'double edged sword' that provokes both mitigating and aggravating consequences (de Kogel & Westgeest, 2015). Muir (2019) for example, found that participants found the information reduced culpability but increased risk. Fuss et al (2015) likewise saw significant reductions in legal responsibility that did not translate into shorter sentences. Path analysis by Cheung & Heine (2015) also highlighted the opposing nature of genetic attributions. Relatively little research has tried to tease these potentially opposing implications apart, but one exception is Allen et al (2019, p. 12) which tried to distinguish between the two hypotheses by using a more sensitive dependent measure. In addition to the standard experimental paradigm of asking participants to indicate a sentence, Allen et al also allowed participants to order a period of hospital treatment. They hypothesised that individuals perceived as less blameworthy and therefore deserving of shorter sentences, might simultaneously be perceived as higher risk and therefore deserving of longer periods of hospital treatment. As predicted, results showed that where individuals were described as suffering from a neurobiological disorder, participants both imposed shorter sentences of imprisonment and longer periods of hospital treatment. Nonetheless, this research left open the question whether the longer periods of hospital treatment had other explanations, such as a belief that the individual's condition was more treatable.

Related to the possible dual nature of the information is the issue of other factors, such as psychopathy, confounding research. Given the uncertainty surrounding exactly what it might be about genotype and environmental information that might mitigate (or aggravate), it remains open that other information disclosed to participants might affect their determinations. References to psychopathy are particularly problematic given the emotive nature of this condition and its known aggravating effect. Research has repeatedly demonstrated that psychopathy is invariably treated as increasing dangerousness and lengthening sentences (Blume et al., 2000, p. 404; S. E. Kelley et al., 2019; Sandys et al., 2009). Psychopathy is such a live issue that participants will often infer it from the information provided, even if it is not disclosed explicitly (Truong et al., 2021). Despite this,

it is relatively common for experimental research examining the effect of propensity information to describe defendants in both control and test conditions as psychopaths (Aspinwall et al., 2012; Kim et al., 2015; Truong et al., 2021). Even where defendants are not described as psychopaths, control conditions are often other types of causal information rather than no causal information at all (Blakey & Kremsmayer, 2018; Gurley & Marcus, 2008). Given the possibility that mitigating implications arise due to the simple existence of a plausible causal explanation, this may be problematic.

In order to distinguish whether propensity information in the MAOA × environment interaction has double-edged effects or no effects, we designed our experimental paradigms to control for the potential opposing effects in two ways. In one way, we explicitly advised all participants that the putative accused was of a higher risk due to his MAOA genotype and the fact that he was abused as a child. Thus, participants in all conditions were shown identical information about the risk posed, thereby controlling for risk. Participants in the test group were additionally advised that the increased risk was because of his MAOA genotype and he was abused as a child. If there was a mitigating effect once the increased risk had been controlled for, we predicted that this would be evident in shorter sentences imposed by the test group. The other way we sought to control for risk was through testing participants across two criminal justice contexts, sentencing and parole. Whereas both blame and risk are appropriate considerations at sentencing, at parole the main consideration is risk. Thus if there was a mitigating effect due to reduced blameworthiness in the test condition, we would expect to see this in the sentencing context but not in the parole context. Finally, to address the confounding effect of increased risk that might be associated with explicit or implicit references to psychopathy or other indications of propensity to offend, we used pared down vignettes, which communicated the bare minimum of necessary information to participants.

## 6.2 STUDY 14

In Study 14, looking at the effects of explanations of propensity to offend while

controlling for risk, the advice used to communicate and control for the increased risk was only verbal. Participants in all conditions were advised that the individual was more likely to react to perceived provocation than average, and correspondingly, was much more likely to be convicted of a violent offence. Participants in the control condition were given no further information, whereas participants in the test condition were advised that this increased risk was explained by the accused's MAOA genotype and because he was abused as a child. All participants were asked to undertake two tasks in a randomised order: to review sentence length in the light of the information provided and to review the risk level in a parole context.

We predicted that in the sentencing context, lengths of imprisonment imposed in the test condition would be lower than in the control condition. We also predicted that in the parole context, there would be no difference between the assessed risk posed by the prisoner.

## 6.2.1 Method

### 6.2.1.1 Participants

All participants were recruited online via Prolific.co. Participants were 102 residents of England and Wales aged between 22 and 71 (thus meeting the eligibility criteria to sit as a lay magistrate in England and Wales). The sample size chosen had 80% power to detect an effect size of w = 0.28 in a $\chi 2$ test and d = 0.57 in a paired t-test (Faul et al., 2007). The mean age was 37; male 39%, female 53%, other 8%; employed full time 48%, part-time 23%, other 29%; student 18%. They were remunerated for their time (£0.50). Participants who participated in one experiment reported here were automatically excluded from the subsequent experiments.

### 6.2.1.2 Design and Materials

Participants undertook two tasks (in a random order) concerning individuals in the

criminal justice system described as posing a higher than average risk. All participants were advised, for each task, that a forensic psychiatric report showed that the individual was much more likely to react to perceived provocation than the average man and, correspondingly, was much more likely to be convicted of a violent offence. Participants suggested a criminal sentence for a convicted accused and they assessed the risk of a prisoner committing a violent offence if released on parole. For both tasks, participants were either given no causal explanation about why the individuals in the scenarios posed a higher risk, or they were told that the higher risk was because the individual had a variant of the MAOA genotype and he was abused as a child. Therefore participants either saw no causal explanation for both tasks, or the same causal explanation for both. The names given to the individuals in the task were randomised (either 'James Worth' or 'Peter Taylor') with one name being used for the sentencing task and the other for the parole task.

For the sentencing task, participants were told that the individual was a 22-year-old man who had been found guilty of assault and that the Magistrates' Court would decide what length sentence of imprisonment he would receive. For the parole risk assessment, participants were told that the individual was a 22-year-old man who had been sentenced to a term of imprisonment for public protection after being convicted of two violent offences, that he had served his minimum sentence term, and the Parole Board would assess the risk he posed of committing a future violent offence to decide whether it was safe for him to be released.

All experiments were programmed using Qualtrics. Both tasks followed a similar format, with participants being asked three questions. For the sentencing task, participants were first asked whether the information in the forensic psychiatric report was relevant to sentence. If they answered negatively, they were asked to explain their reasoning and were asked no further questions about sentencing. If they answered positively, they were secondly asked whether they thought the information was aggravating or mitigating. Thirdly, they were advised that the individual would ordinarily receive a sentence of 6 months, and they were invited to revise the sentence in the light of the information in the forensic psychiatric report.

Here they were able to select a sentence in the range of 0 to 12 months (in increments of 0.1 months). Correspondingly, in the parole context, participants were first asked whether the forensic psychiatric report was relevant to parole. If they answered negatively, they were asked to explain their reasoning and were asked no further questions about parole. If they answered positively, they were secondly asked whether the information increased or decreased the risk that the individual posed. Thirdly, they were advised that the individual's risk would ordinarily be assessed at 50 on a continuous scale from 0 to 100 (in increments of 1) where 0 was no risk and 100 was the highest risk, and were invited to revise their risk assessment in light of the information in the forensic psychiatric report. For reasons of external validity, and to increase engagement with the task, participants were also asked to explain their reasoning for all judgments, but this qualitative data was not analysed for the purpose of the present report.

#### 6.2.1.3 Procedure

After following a link from Prolific.co, participants first provided informed consent to participate in the experiment. In the experimental instructions they were next advised that recent research had found that some men are more likely to react to perceived provocation and correspondingly to be convicted of violent crimes than the average man. Participants were told that we were interested in their views about the relevance of this information to criminal justice and that they would be asked for their views about two hypothetical criminal justice scenarios.

Participants who completed the survey were thanked, debriefed, and redirected to Prolific.co to record their successful completion. Those who exceeded the 30-minute maximum participation time were automatically excluded by Prolific.ac. The average participation time was 6 minutes.

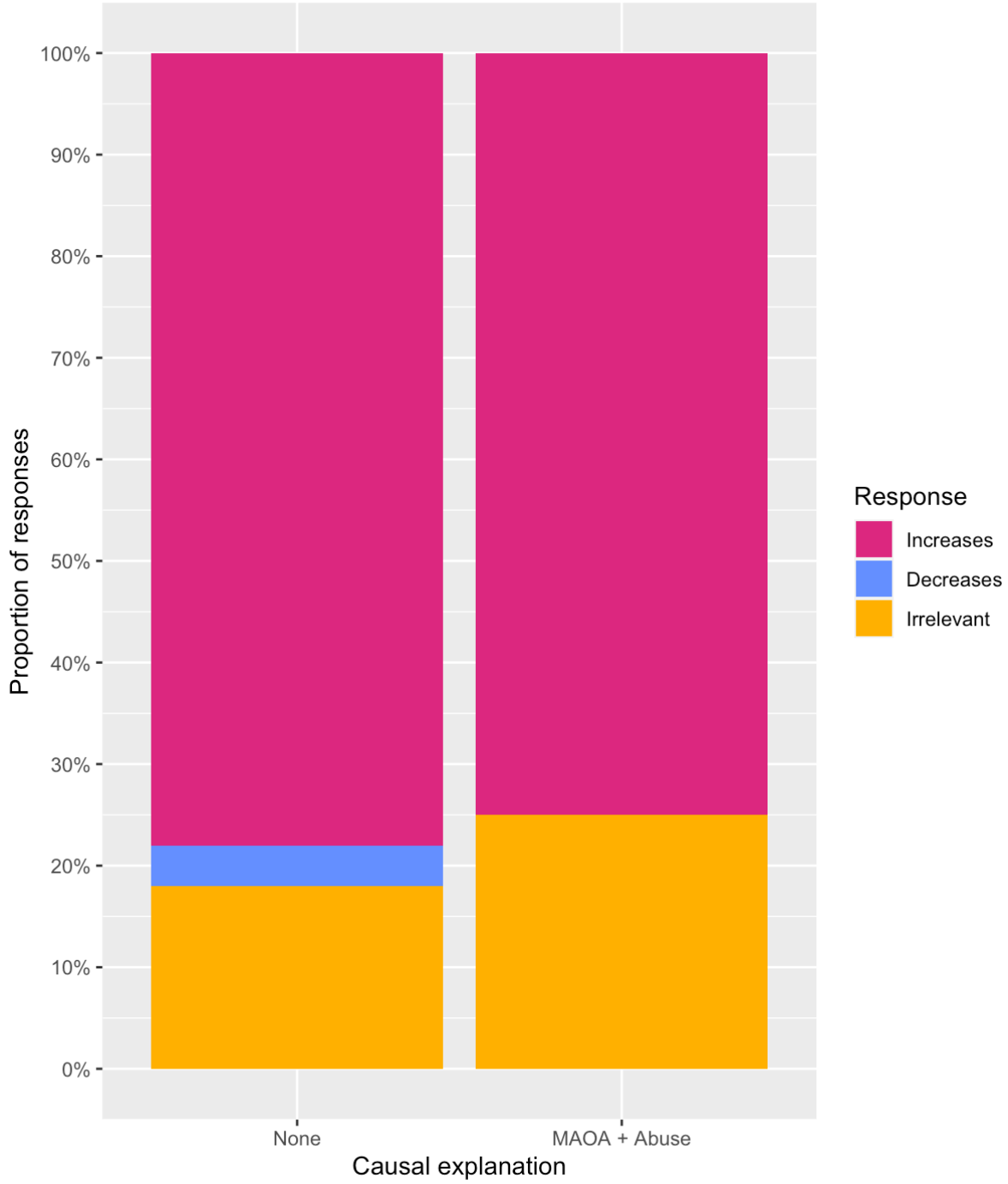
Ethical approval to conduct all experiments was provided by the Ethics Chair for

UCL's Speech Hearing and Phonetic Sciences Research Department (project ID No: SHaPS-2015-AH-017).

### 6.2.2 Results

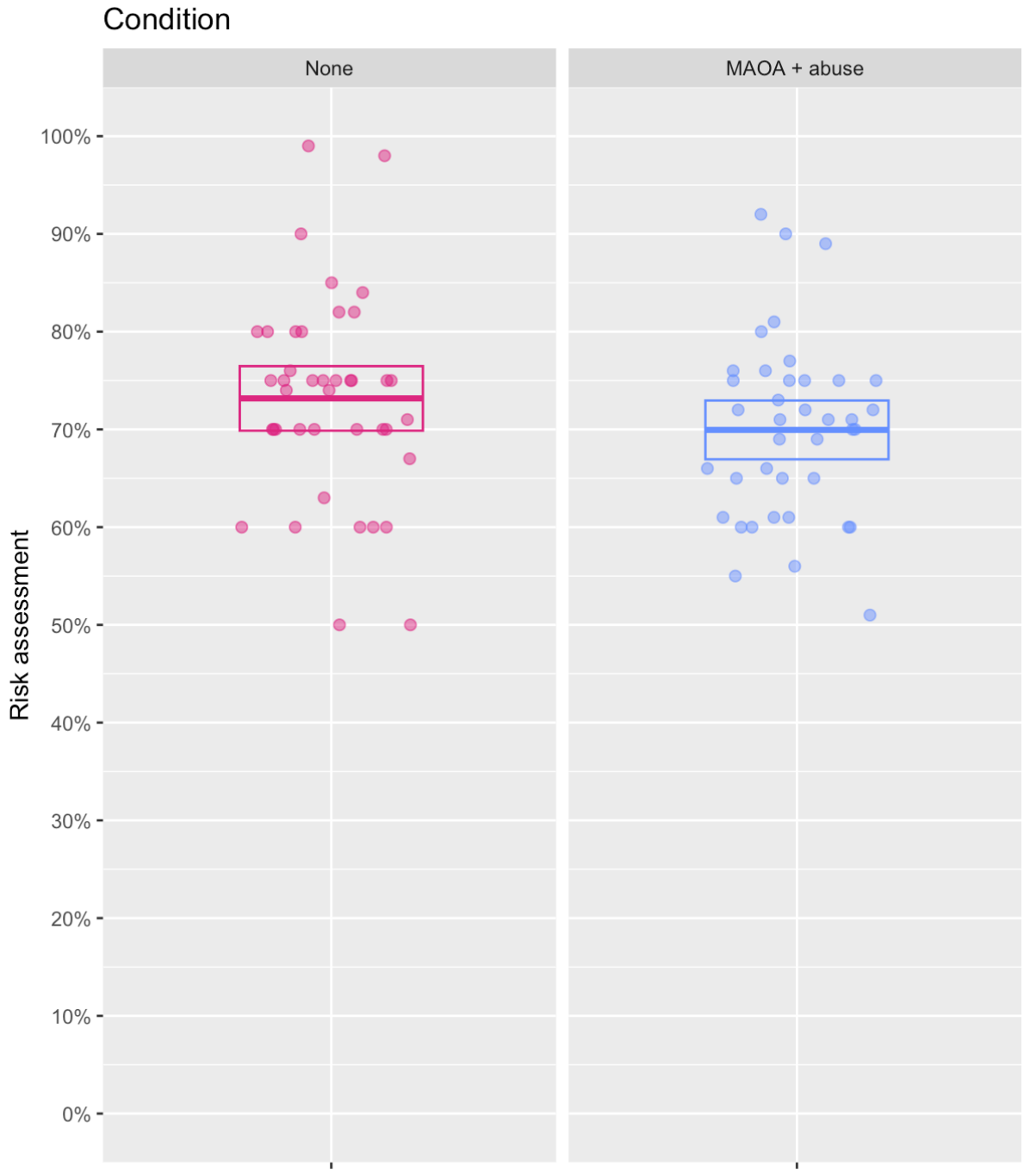
Overall, fewer participants thought that the psychiatric report was relevant to the sentence context (51%) than the parole context (78%). In the parole context, we coded participants' responses as 'irrelevant', 'increases risk', and 'decreases risk'. The distribution of responses did not differ according to whether it was accompanied by the causal explanation or not (see Figure 17): $\chi 2$ (2, $N$=102) = 2.69, $p$ = 0.26.

*Figure 17. Proportion of participants assessing parole report as increasing or decreasing risk or being irrelevant, by causal explanation for Study 14.*



Also in the parole context, amongst those treating the information as relevant, there was no difference in the actual risk assessments (causal = 69.9%; non-causal = 73.2%), mean difference = 3.3%, CI = [-1.2%, 7.6%], $t$ (78) = 1.46, $p$ = .15, $d$ =0.33. See Figure 18.

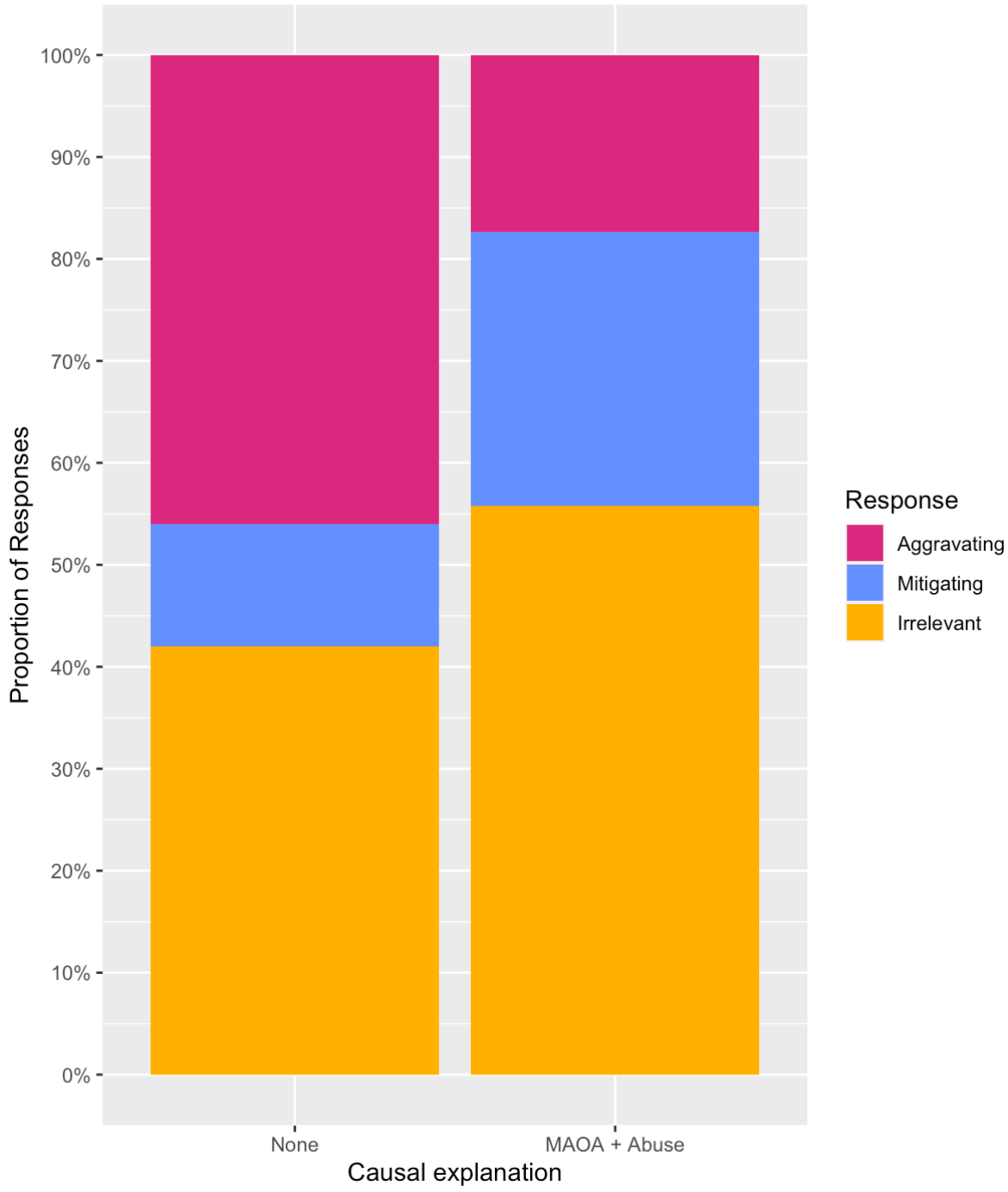
*Figure 18. Parole risk assessments by condition for Study 14.*




In the sentence context we similarly coded participant's responses as 'irrelevant', 'aggravating', and 'mitigating' and here, by contrast, the responses did differ: $\chi 2$ (2, $N$=102)

= 10.57, $p$ = 0.005, Cramer's V = 0.32, see Figure 19. Whilst there was no difference in the proportion of participants perceiving the report to be irrelevant, $\chi 2$ (1, $N$=102) = 1.42, $p$ = .23, amongst those who treated the information as relevant, participants were much more likely to treat the information as mitigating if they received the causal explanation (61%) than if they did not (21%), difference = 40%, CI = [16%, 64%], $\chi 2$ (1, $n$=52) = 7.13, $p$ < .01, Cramer's $V$=0.37.

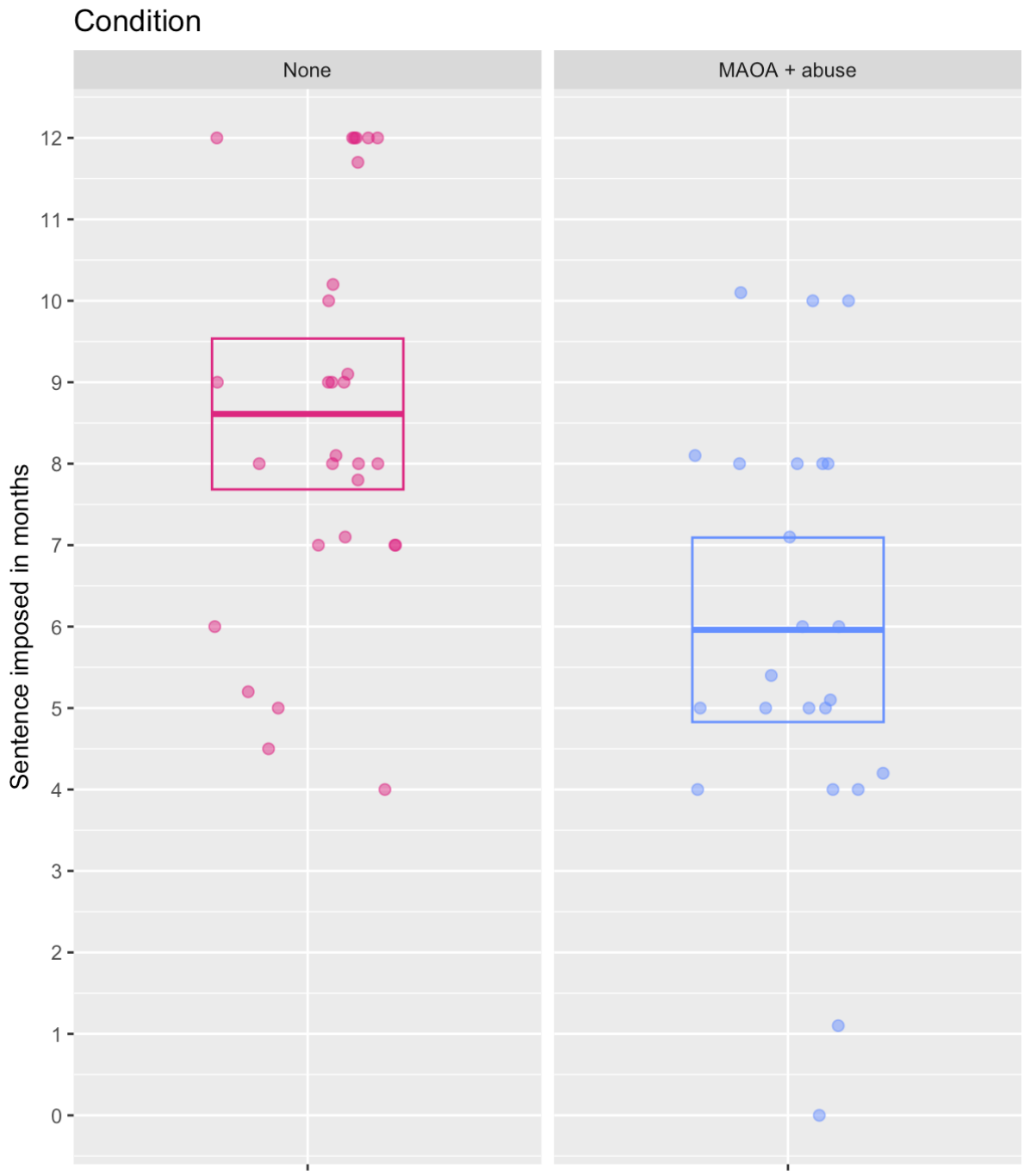
*Figure 19. Responses to information in psychiatric report by condition for Study 14.*




This translated into a much shorter suggested sentence where the causal explanation was included (5.96 months) than where it was not (8.61 months): difference between sample

means (mean difference) = 2.65 months, CI = [1.24, 4.06], $t(50)= 3.77$, $p < .0001$, $d=1.053$ (see Figure 20).

*Figure 20. Sentence imposed by condition in Study 14.*



### 6.2.3 Discussion

The results of Study 14 were as predicted, but also highlighted a number of interesting collateral findings. One such finding was the relatively high proportion of participants (roughly half) who considered the information about the increased risk irrelevant to sentence, and that this did not differ between the control and test conditions. By contrast, a considerably smaller proportion of participants (roughly a fifth) thought the information irrelevant to parole, which again did not differ between conditions. Given that participants were asked to revise sentence lengths from the 6 months that they were told would ordinarily be imposed, it appeared that the information that the accused was a greater risk was treated as quite an aggravating factor in the abstract, given that participants in the control condition who considered the information relevant imposed a mean sentence of 8.61 months, almost 50% longer than the 6 month starting point. Yet the explanation for the increased risk seemed to have a significant mitigating effect once the increased risk was controlled for, with participants in the test condition who considered the information relevant imposing sentences 2.65 months shorter than in the control condition. What was quite illuminating was that the final mean sentence imposed by participants in the test condition was 5.96 months, very close to the original 6 month starting sentence. This was consistent with the double-edged sword hypothesis and suggested that had the increased risk not been controlled for, it would have been difficult to distinguish between the aggravating effect of risk and the mitigating effect of the explanation compared to a lack of effect of either.

By contrast, in the parole context where the primary consideration is the future risk of reoffending rather than other matters such as blameworthiness, there was no difference between the conditions. As well as the overwhelming majority of participants considering the information to be relevant to the parole decision, almost all treated the information as increasing risk. Most noteworthy was the fact that here there was no significant difference between the conditions, suggesting that to the extent that explanations of the increased risk affect decisions in a criminal justice context, they affect considerations other than risk.

## 6.3 STUDY 15

What remained striking about Study 14, notwithstanding our predictions, was the drastically shorter sentence imposed on offenders by participants who treated the information as relevant when the risk that the individual posed was held constant across the two scenarios. In both scenarios there must have been some causal explanation for the offending, with the difference being that in the control condition no explanation was given whereas in the test condition a plausible explanation was given. It seemed surprising that this information was sufficient to influence behaviour so considerably. Given the scarcity of information about the risk posed by the accused, it was possible that participants who read about the MAOA genotype and childhood maltreatment were using this causal information to quantify the risk posed by the accused. Consequently, in Experiment 15, we focussed in on the sentencing context only and tested whether the effect of the causal information still obtained when concrete and identical quantitative information about risk level was visually communicated to participants in the form of an infographic.

### 6.3.1 Method

#### 6.3.1.1 Participants

The number of participants per cell was doubled from Experiment 14, in acknowledgement of the fact that half the participants previously perceived the psychiatric report as irrelevant to the sentence task. All participants were recruited online using prolific.ac. Participants were 400 residents of England and Wales aged between 20 and 75 with a mean age of 37.4; male 43.2%, female 54.8%%, other 2.0%; employed full time 50.5%, part-time 26.8%, other 22.7%; Student 23.4%. They were remunerated for their time

(£0.50).

### 6.3.1.2 Design, Materials, and Procedure

In Study 15, we based the quantitative risk information (provided to half of the participants in Study 14) on the evidence from Caspi et al.'s (2002) study of a large cohort of New Zealand males. We also provided an infographic that displayed this risk in a visual form. This time, we focused solely on the sentencing context. Thus, in addition to the previous variable of whether participants were given information about the causal explanation for the increased risk, we added a second variable of whether the participant was shown the quantitative statistical information detailing the precise level of increased risk that the accused posed. Thus, we employed a 2 (causal explanation present or absent) × 2 (quantitative risk information present or absent) between-participants design.

Participants receiving quantitative risk information were told at the outset that a forensic psychiatric report showed that the individual belonged to a higher risk group of 12% of the population, that men in the general population would on average commit 0.4 violent crimes before age 26, and that of these, men in the higher risk group would on average commit 1.5 violent crimes before age 26, whereas men in the rest of the population will on average commit 0.25 violent crimes before age 26. Those in this group were also shown the infographic shown at Figure 21 which conveyed the same information visually. This information was shown at the point that participants assessed whether the information provided was relevant to sentence.

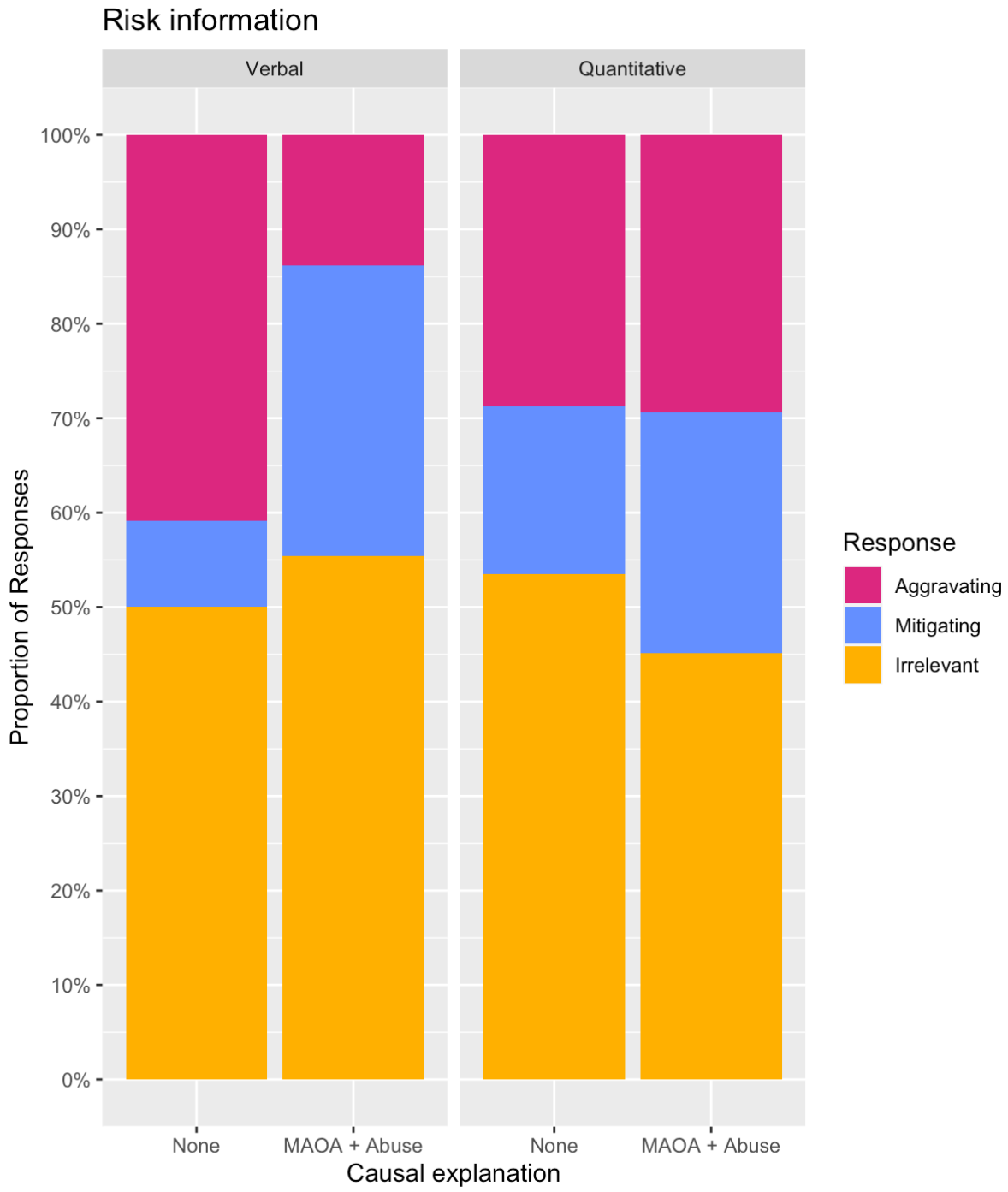
*Figure 21. Risk infographic used in Studies 15, 16, and 17.*


### 6.3.2 Results

As with Study 14, we coded the responses to the sentence task as 'irrelevant', 'aggravating', and 'mitigating'. The distribution of responses again differed according to whether the report was accompanied by the causal explanation or not: $\chi 2$ (2, $N$=402) = 16.21,

$p < 0.001$, Cramer's V = 0.20. There was no difference between participants who saw the quantitative risk information ($\chi 2$ (2, $n$=203)= 2.11, $p$ = .35, Cramer's $V$ = 0.10) whereas there was a difference for those who only saw the verbal risk information ($\chi 2$ (2, $n$=199)= 25.05, $p$ < .001, Cramer's $V$ = 0.35), see Figure 21. As with Study 14, about half (49%) of participants thought the forensic psychiatric report relevant to sentence, and there was no difference according to whether the causal explanation was included or not: difference = 2%, CI = [ -8%, 11%], $\chi 2$ (1, N=402) = 0.041, p = .84; or by whether they had seen the quantitative risk information: difference = 4%, CI = [-6%, 13%], $\chi 2$ (1, N=402) = 0.36, p = .55. Also consistent with Study 14, of participants who considered the report relevant, those presented with the causal explanation for the accused's risk were much more likely to treat this as mitigating (57%) compared to those receiving no causal explanation (28%). A logistic regression analysis showed that the effect of the causal explanation was significant: b(Odds) = 0.10, CI = [0.16, 0.54], $\chi 2$ (1, n=196) = 16.47, p < .001, McFadden's pseudo R2 =0.06 (see Table S1). The interaction between the two conditions was significant: b(Odds) = 7.05, CI = [2.01, 25.08], $\chi 2$ (1, n=196) = 9.91, p< .01, McFadden pseudo R2 = 0.10, suggesting that participants were more likely to treat the causal explanation as mitigating where they had only been advised of the risk verbally compared to if they had also been provided with quantitative information about the risk.

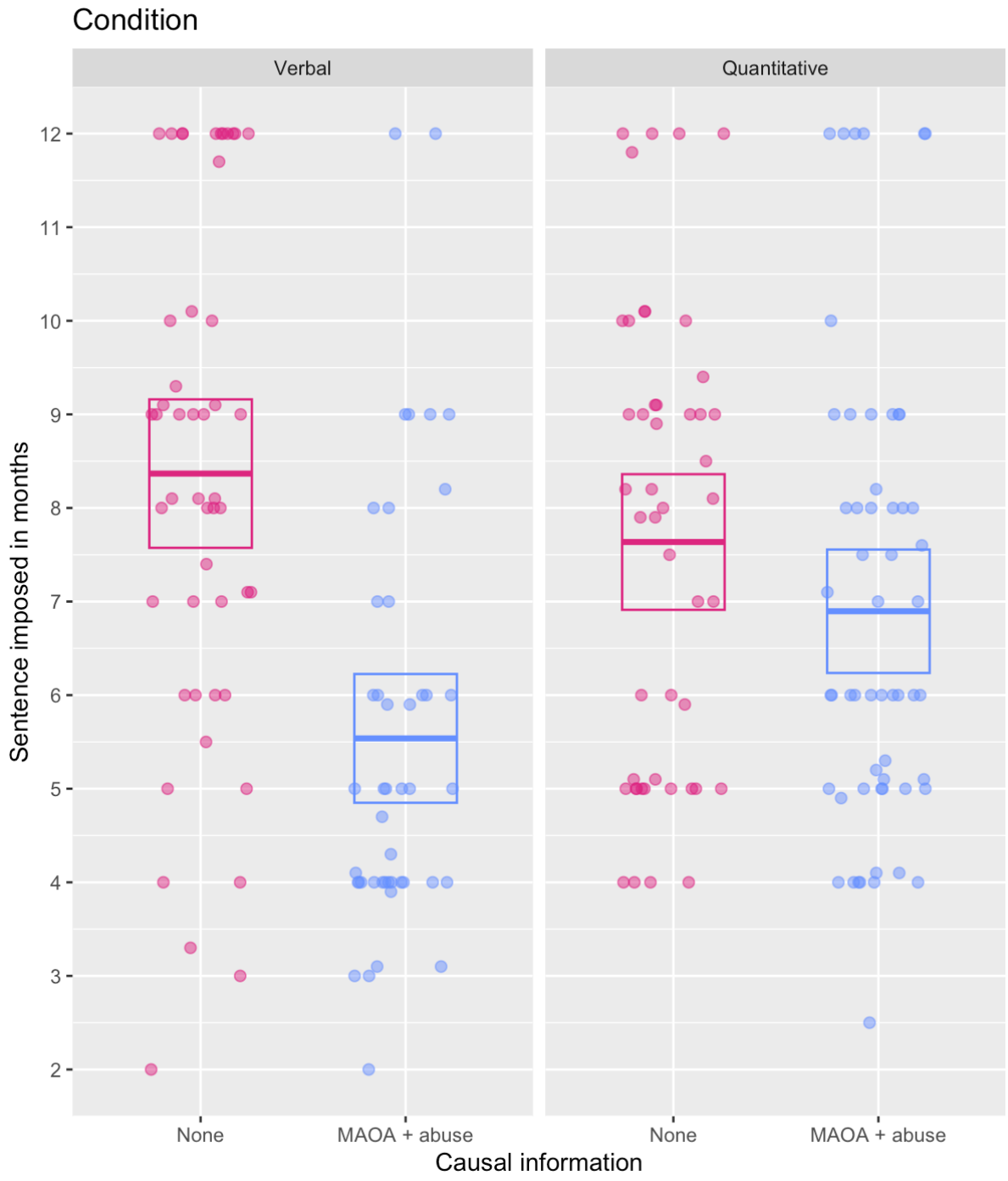
*Figure 22. Participant assessments that the psychiatric report was aggravating or mitigating or irrelevant by causal information and risk information for Study 15.*




The above pattern was reflected in the actual sentences imposed, see Figure 23. As

with the mitigating or aggravating responses, there was a significant causal explanation present / absent × quantitative information present / absent interaction: difference = 2.09 months, CI = [0.68, 3.50], F (1, 193) = 8.52, p <.001, η2=0.04, suggesting that sentences were moderated by the quantitative information such that sentences were slightly shorter where no causal information was provided and sentences were slightly longer where causal information was provided.

*Figure 23. Sentences imposed by causal information and risk information for Study 15.*



### 6.3.3 Discussion

Study 15 replicated the mitigating effect of the causal explanation on sentences seen in Study 14 for those participants given the verbal risk information. The size of the effect was, however, reduced when provided with the quantitative risk information. Study 15 presented the quantitative risk information only at the first stage of the experiment where participants indicated whether the information was relevant to sentence, and therefore not where participants indicated whether they thought the information was mitigating or aggravating, and imposed a sentence. Given the relative complexity of the information and the infographic, we considered that there was a risk that this design imposed an unnecessary cognitive load on participants that we had not intended. For this reason, we decided to rerun the study with the quantitative risk information displayed at every stage, to avoid the potential for participant confusion to be affecting the results.

## 6.4 STUDY 16

Study 16 was identical to Study 15, other than that for participants shown the quantitative information and accompanying infographic conveying the risk, this was shown at each stage, rather than only at the outset as with Study 15.

### 6.4.1 Method

#### 6.4.1.1 Participants

As with Study 15, the number of participants per cell was doubled from Study 14, in acknowledgement of the fact that we expected that around half of participants would consider the information irrelevant to sentence. Participants were 402 residents of England and Wales.

Participants were aged between 20 and 74 with a mean age of 38; male 41%, female 58%, other 1%; employed full time 49%, part time 23%, other 28%; student 22%. They were remunerated for their time (£0.60). Average participation time was 4 minutes.

#### 6.4.1.2 Design, Materials, and Procedure

The design, materials, and procedure for Study 16 were identical to Study 15 with one exception. In Study 16, participants receiving the quantitative risk information were able to view it throughout the experiment rather than only at the start of the experiment.

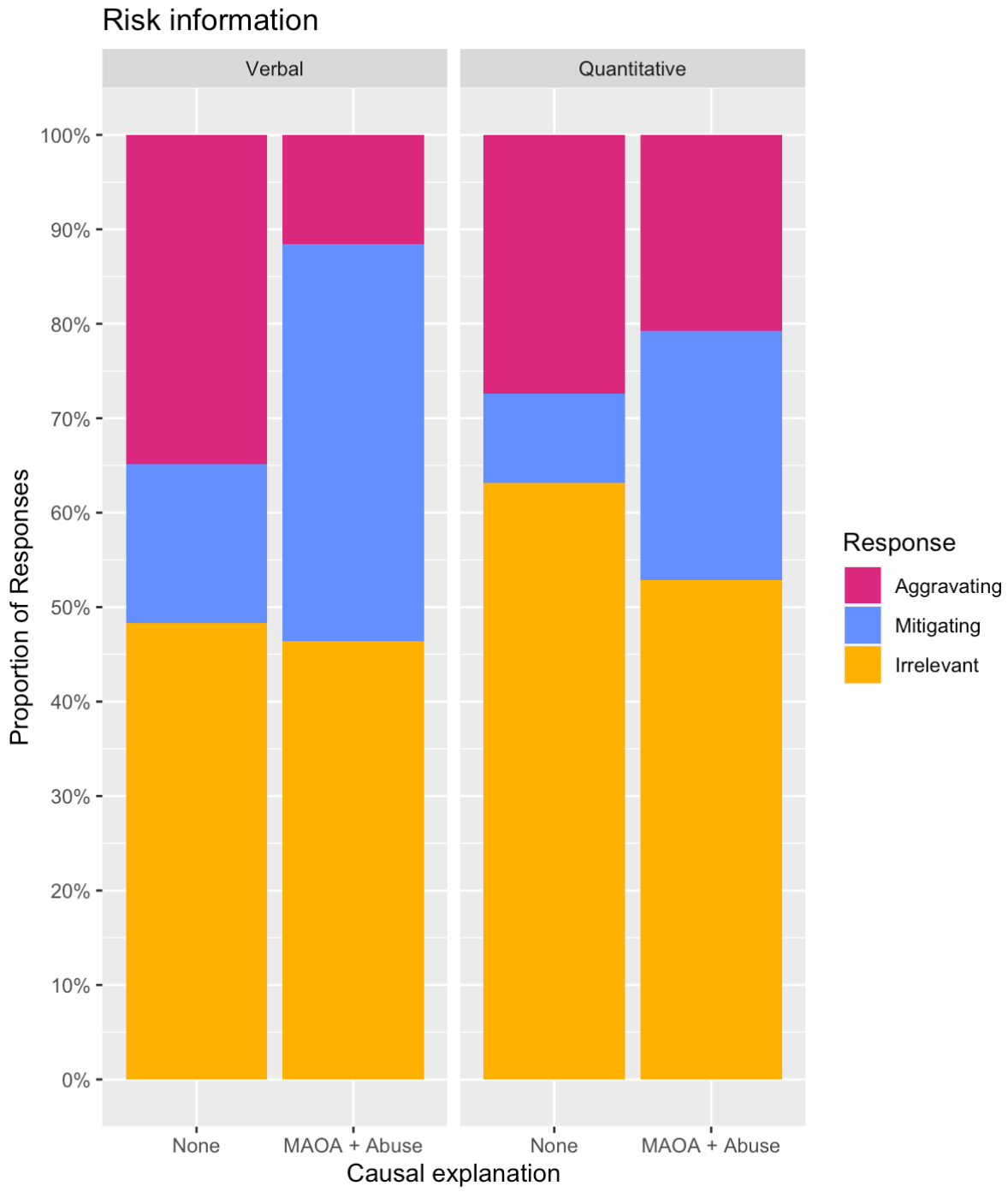
### 6.4.2 Results

As with previous studies, we coded the responses to the sentence task as 'irrelevant', 'aggravating', and 'mitigating'. As before, the distribution of the three responses in this sentencing context differed according to whether the report was accompanied by the causal explanation or not: $\chi 2$ (2, $N$=402)= 28.98, $p < .001$, Cramer's V = 0.27. This effect was observed both for those who received quantitative risk information ($\chi 2$ (2, $n$=201)= 9.65, $p < .01$, Cramer's $V = 0.22$), and for those who did not ($\chi 2$ (2, $n$=201)= 22.39, $p < .001$, Cramer's $V = 0.33$).

A similar proportion to Study 14 thought the forensic psychiatric report relevant to sentence (48%), and again the perceived relevance of the report did not differ according to whether it was accompanied by the causal explanation or not: difference = 6%, CI = [-3%, 16%], $\chi 2$ (1, $N$=402) = 1.410, $p = .24$. Those who received the quantitative risk information were slightly less likely to consider the report relevant (42%) than those who had not (53%), difference = 11%, CI = [0.7%, 20%], $\chi 2$ (1, $N$=402)= 3.99, $p = .046$, Cramer's $V = 0.10$. As in Study 14, of those who thought that the report was relevant, participants presented with the

causal explanation were much more likely to treat the risk information as mitigating (68%) compared to those receiving no causal explanation (30%; see Figure 24). A logistic regression confirmed that the effect of the causal explanation was significant: OR = 0.18, CI = [0.095, 0.34], $\chi 2$ (1, $n$= 191) = 29.81, $p$<.001, McFadden's pseudo $R2$=0.108 (see Table 1). The effect of the quantitative risk information was also significant, but with a much smaller effect size: OR = 2.13, CI = [1.14, 4.06], $\chi 2$ (1, $n$=191) = 5.59, $p$=.018, McFadden's pseudo $R2$=0.016. Crucially, this time, the interaction term was not significant: OR = 2.03, CI = [0.56, 7.33], $\chi 2$ (1, $n$=191) =1.168, $p$=.28, indicating that the effect of the provision of the causal explanation was unaffected by the inclusion of the quantitative risk information.

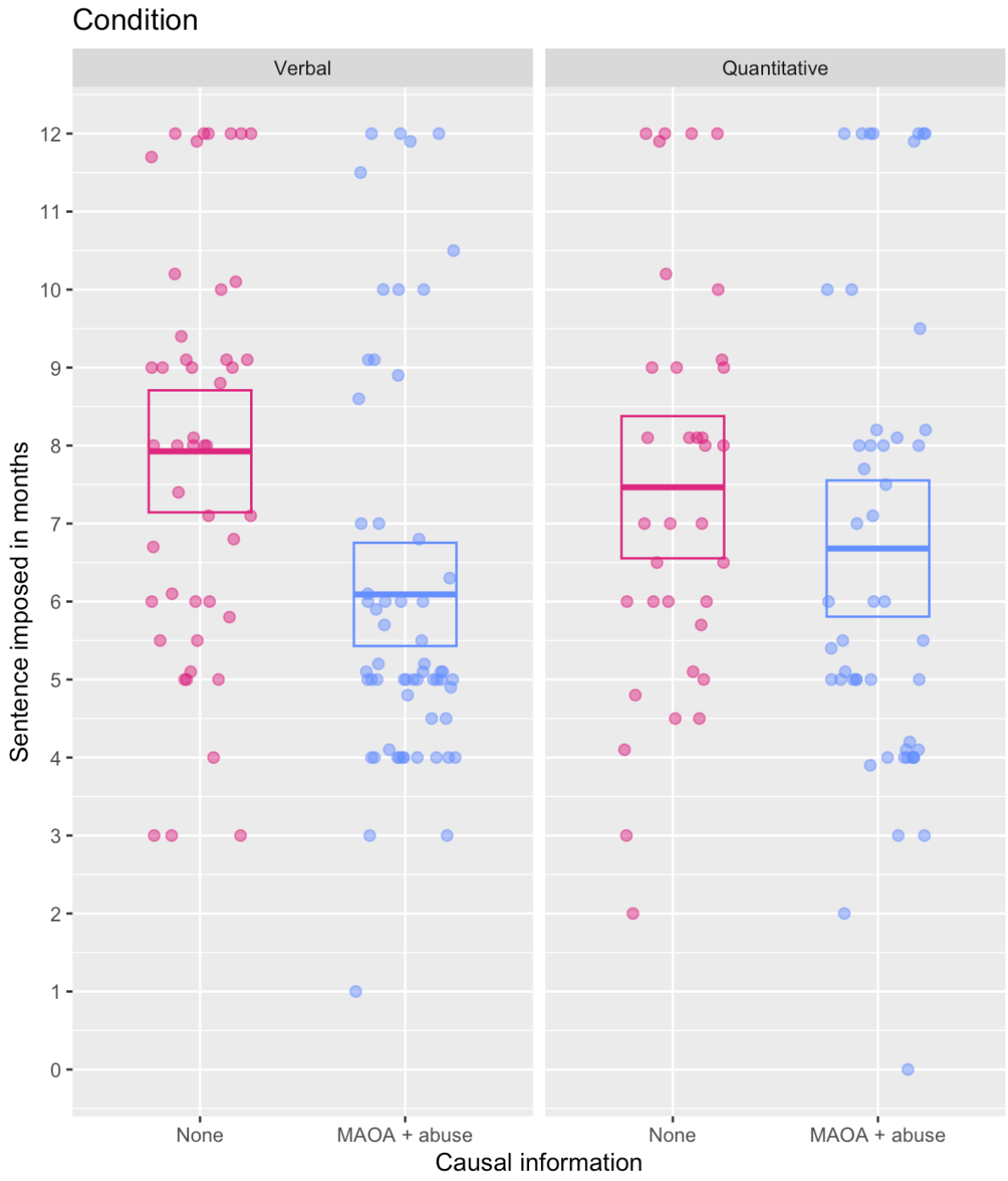
*Figure 24. Participant assessments that the psychiatric report was aggravating or mitigating or irrelevant by causal information and risk information for Study 16.*




These figures translated into the same pattern of effects on sentence as Study 15. Of

those who thought that the report was relevant, those shown the causal explanation about MAOA genotype and childhood abuse imposed a shorter mean sentence than those not provided with any causal explanation, despite seeing exactly the same level of risk: mean difference = 1.83 months, CI = [0.78, 2.89], $F$ (1, 187) = 11.65, $p$ <.001, η2= 0.06 (see Figure 25). There was no significant interaction: CI = [-0.54, 2.64], $F$(1,187) = 1.69, $p$=.20, η2= 0.008 and no main effect of the quantitative risk information: mean difference = 0.46 months, CI = [-0.75, 1.67], $F$(1,187) = 0.13, $p$=.71, η2= 0.001).

*Figure 25. Sentences imposed by causal information and risk information for Study 16.*



### 6.4.3 Discussion

Study 16 confirmed the original findings of Study 14 and also indicated that the effect still manifested itself where participants were given very precise quantitative information about the increased risk posed by the accused accompanied by an infographic, rather than a more uncertain verbal indication of increased risk. Contrary to the findings in Study 15, it was also clear that the effect was very similar regardless of whether participants were given more vague or more precise information, suggesting that the apparent moderating effects seen in Study 15 were due to not continuing to show the more precise information when participants were deciding whether the information was mitigating or aggravating, or when they were imposing a sentence.

## 6.5 STUDY 17

In our final study, we sought to replicate our previous research in both the sentence and parole context. In Study 17, all participants were provided with the quantitative information about the precise level of risk posed by the accused including the infographic. This replication was considered important given the slightly different results observed across Studies 15 and 16. We also added a specific question asking all participants (even those who indicated that the information was irrelevant to their decision) about the effect of the forensic psychiatric report on blameworthiness and risk in the sentencing context. For these additional questions, we hypothesised that those told that the accused's higher risk was caused by his MAOA genotype and childhood abuse would rate him as less blameworthy than those in the control group, but that there would be no difference in the assessments of risk.

### 6.5.1 Method

#### 6.5.1.1 Participants

Participants were selected as residents of England and Wales aged between 40 and 65 to be a more representative age range for lay justices and parole board members. The 404 selected participants were aged between 40 and 65 with a mean age of 50; 39% male, 61% female; employed full time 49%, part time 19%, other 32%; 1.5% students. They were remunerated for their time (£0.55). Average participation time was 8 minutes.

#### 6.5.1.2 Design and Materials

Participants again completed both the sentencing task and the parole task. All participants were shown the quantitative information about higher risk including the infographic, as in the quantitative information condition of Experiments 15 and 16 (see Figure 21), with half of them again provided with the additional causal explanation. The order of the sentencing and parole tasks was not randomised: all participants undertook the sentencing task first. This was to facilitate asking additional direct questions about blame and risk after the sentencing task and before the parole task, because asking questions about blame made little sense in the context of a risk assessment. All participants (even those who previously indicated that the information was irrelevant) were asked directly what effect the information in the forensic psychiatric report had on blame, and on risk. For each question, participants were able to select an answer from a 5-point Likert scale with the available choices being: significantly decreases [blame/risk], slightly decreases [blame/risk], no effect / irrelevant, slightly increases [blame/risk], significantly increases [blame/risk]. All other elements of the procedure were identical to Study 14.

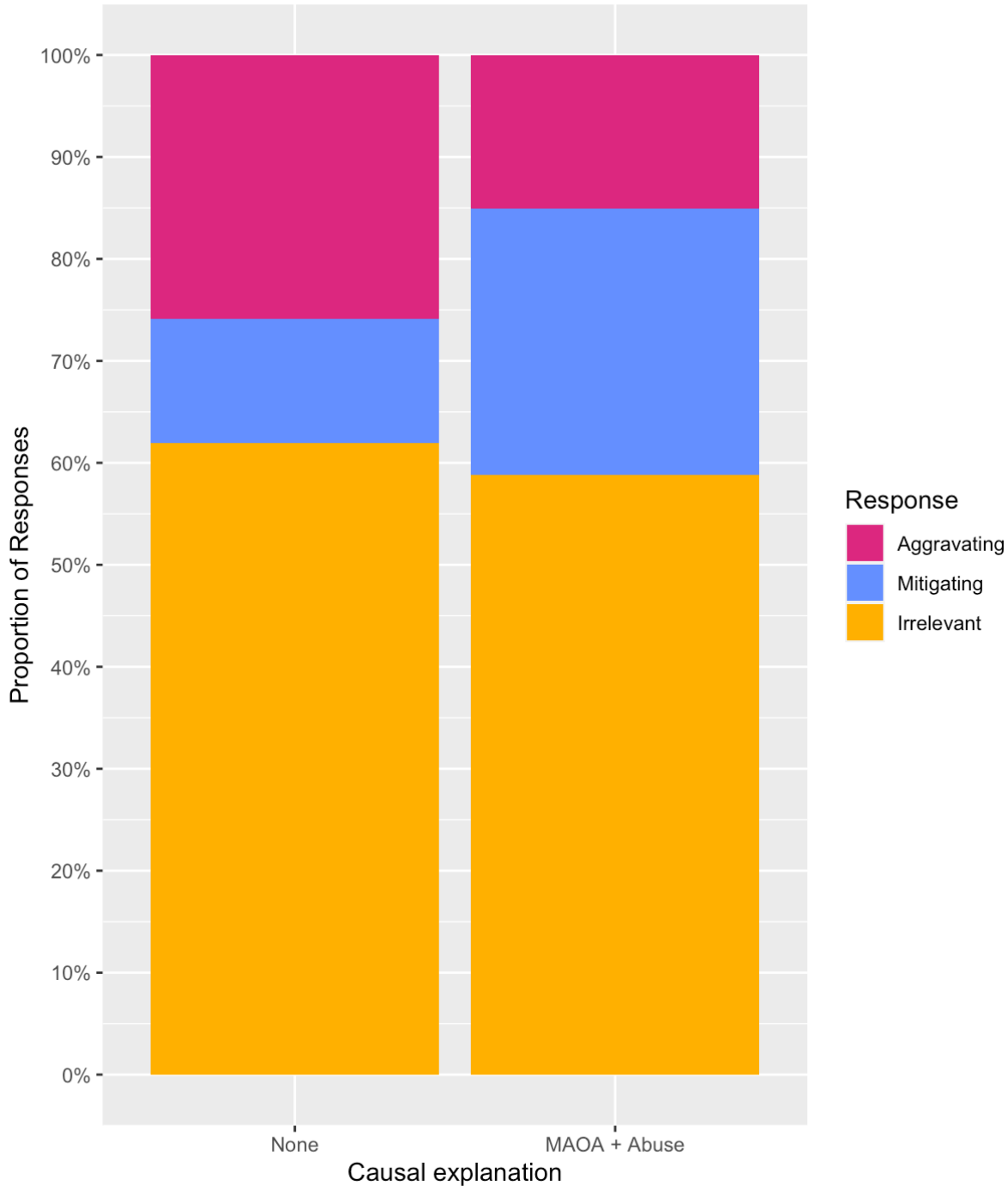
#### 6.5.1.3 Procedure

The Procedure for Experiment 15 was the same as Experiment 13, save that participants always undertook the sentencing exercise first (rather than being in a randomized order with the parole exercise). Immediately after the sentencing exercise, participants were asked the direct questions about the effect of the information provided on blame and risk. Following the direct questions, participants undertook the parole exercise. Average participation time was 8 minutes.

### 6.5.2 Results

Consistent with previous results, the proportion of responses coded as 'irrelevant', 'aggravating', and 'mitigating' differed according to whether participants received the causal explanation or not $\chi 2$ (2, $N$=404) = 16.16, $p < 0.001$. Slightly fewer participants than in previous experiments thought that the information was relevant to sentence (40%), but again the difference by whether they received the causal explanation or not was not significant: mean difference = 4%, CI = [-5%, 14%], $\chi 2$ (1, $N$= 404)= 0.56, $p$ = .45. As before, of participants who considered the information relevant to sentence, those receiving the causal explanation were much more likely to treat the risk information as mitigating than were those who received no causal explanation: difference = 31%, CI = [16%, 45%], $\chi 2$ (1, $n$=150) = 14.52, $p < .001$, Cramer's $V$ = 0.30, see Figure 26.

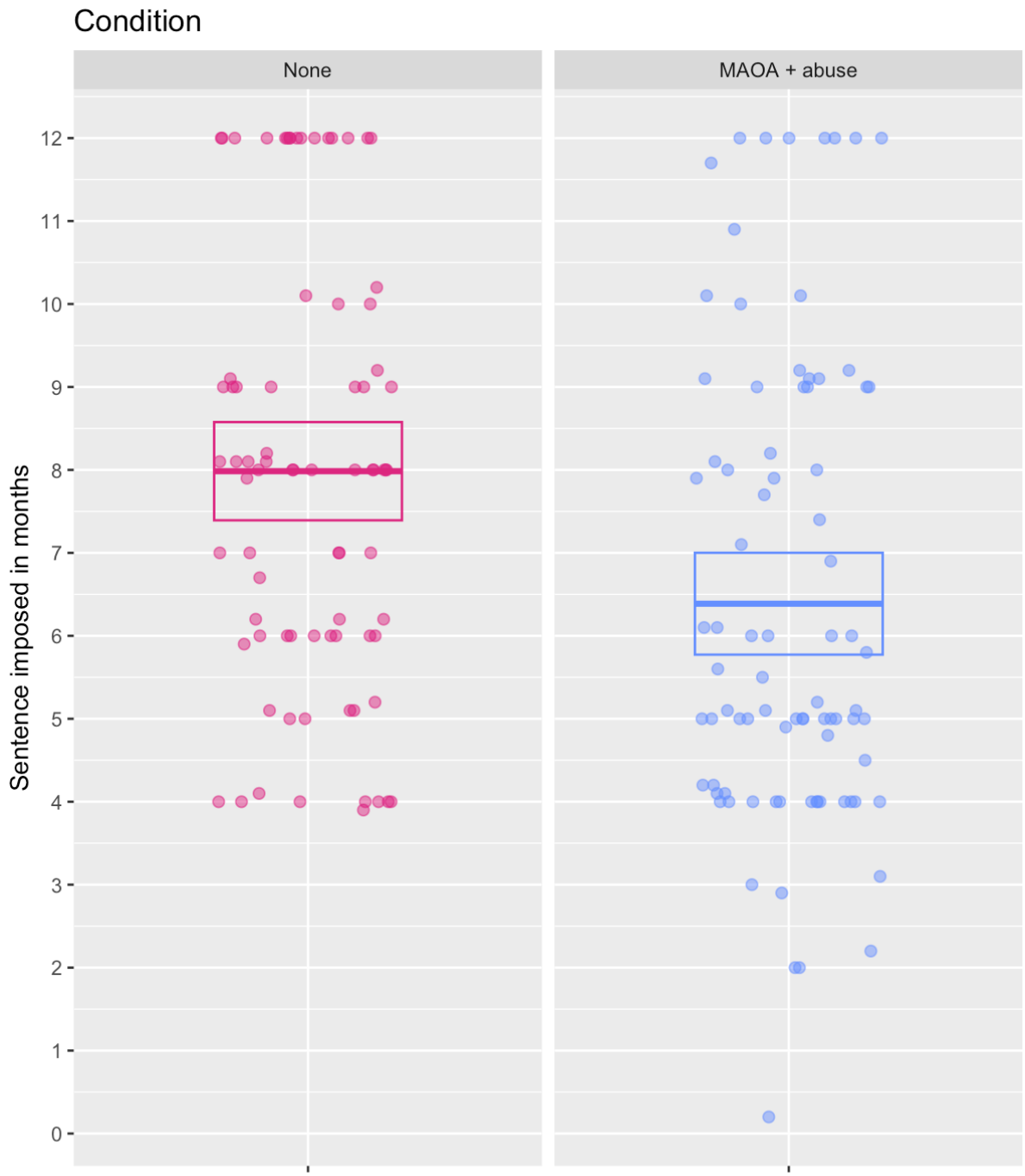
*Figure 26. Responses to information in psychiatric report by condition for Study 17.*




Correspondingly, of those who considered the information relevant, much shorter sentences were indicated by those who received the causal explanation (6.39 months) than

those who did not (7.98 months), mean difference = 1.60 months, CI = [0.75, 2.45], $F(1, 157) = 13.77$ $p<.001$, $\eta 2 = 0.08$ (see Figure 27).

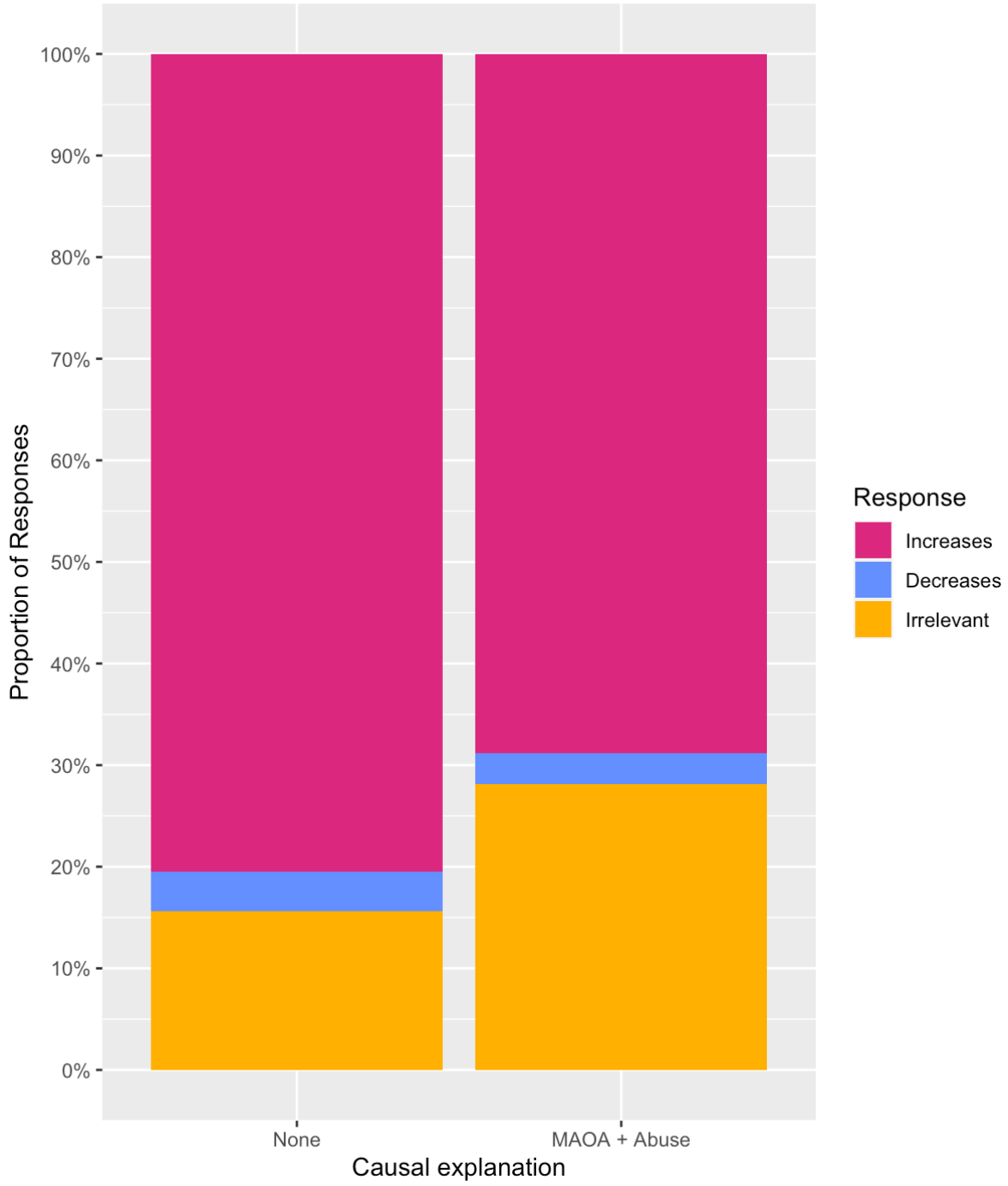
*Figure 27. Sentences imposed by condition for Study 17.*



Responses to the new blameworthiness and risk questions asked of all participants were converted to a -2 to +2 scale for analysis (significantly/slightly increases/decreases / no effect). We undertook a linear mixed effects analysis incorporating an intercept for participant as a random effect. This yielded a significant 'causal explanation present/absent × blameworthiness/risk judgment' interaction: CI = [0.31, 0.71], $F$ (1, 404.65) = 25.06, $p$ < 0.001. Analysis of simple effects confirmed that participants considered that the information increased risk but had little or no effect on blame, save where participants were provided with the causal information, when the information was taken to reduce blame. Thus, participants provided with the causal explanation considered the information in the forensic psychiatric report to reduce blameworthiness significantly more (-0.34) than those not provided with the causal explanation (0.08), difference = 0.42, CI = [0.28, 0.57], $t$(808) = 5.86, $p$ < .001. Also as predicted, no significant difference was observed between participants' risk judgments where causal explanation was provided (1.17) compared to when it was not (1.08), difference = 0.09, CI = [-0.05, 0.23], $t$(808) = 1.21, $p$ = .23.

Coincidentally, exactly the same proportion as in Study 14 (78%) thought the information was relevant to parole. Here, the distribution of responses coded as 'irrelevant', 'increasing', and 'decreasing' differed somewhat according to whether it was accompanied by the causal explanation or not: $\chi 2$ (2, $N$=404) = 9.34, $p$ = 0.009, see Figure 28. Participants were slightly less likely to treat the risk information as relevant when provided with the causal explanation (71%) than when not (84%), difference = 13%, CI = [5%, 21%], $\chi 2$ (1, $N$=404)= 8.59, $p$ = .0034, Cramer's $V$ = 0.146.

*Figure 28. Proportion of participants assessing parole report as increasing or decreasing risk or being irrelevant, by causal explanation for Study 17.*




However, as in Study 14, participants who considered the information relevant to their

decision overwhelmingly treated the risk information as increasing risk (95.6%) and there was no difference between the experimental conditions, $\chi 2$ (1, $n$=306) < 0.001 , $p$ = 1. Correspondingly, there was also no difference in quantitative risk assessments made by participants, with those seeing the causal explanation assessing the risk posed at 72.0% and those not seeing the causal explanation assessing the risk as 71.9%, see Figure 29.

*Figure 29. Parole risk assessments by condition for Study 17.*

### 6.5.3 Discussion

Study 17 used the quantitative information about the risk posed throughout, rather than the less specific verbal descriptions used in Studies 14, 15, and 16. It was also accompanied, as before, by the infographic. As such, while Study 16 confirmed that there was no real difference in effect between the qualitative and verbal information about risk, the specificity of the quantitative information minimised the possibility that any participants were using the causal information to assess any lingering uncertainty about the risk posed by the individuals in the scenarios. We therefore were reasonably confident that we had controlled for the increased risk posed by the individuals due to their circumstances.

From the sentencing perspective, Study 17 confirmed and replicated the pattern seen previously in all experiments whereby those participants deeming the information relevant treated it as aggravating when only aware of the increased risk posed by the individual, but this aggravation was effectively cancelled out by the mitigating effects when the causal information accounting for this increased risk was presented. The new explicit questions introduced in this study about effects on blame and risk seemed to be very consistent with this pattern. Participants treated the information as increasing risk and this did not differ by whether they were provided with the causal explanation or not. By contrast, there was a significant difference between assessments of effect on blame, with participants only provided with the information about increased risk, treating this as having little or no effect on blame, or being irrelevant to blame. Those given the causal explanation treated it as reducing blame.

Of additional interest was the relatively high proportion of participants who again indicated that the further material was irrelevant to sentence regardless of its content. In Study 17 reached its highest point of around ⅔ of participants. While not the focus of this study, this seems to be a phenomenon worthy of further investigation and explanation.

The parole experiments that were also replicated showed the same pattern as in Study

14. A high proportion of participants treated the additional information as relevant to parole, though this time there was a modest statistical difference between the conditions, with participants given the causal information less likely to treat the information as relevant to their decision. However, of those who used the information, the effect was to increase risk, with no statistical difference between the conditions. This again reinforced the view that once risk was controlled for, there was little or no effect of the causal information, suggesting that to the extent that causal information affected decisions in a criminal justice context, its influence was on factors other than risk.

Overall, the conditions replicated in the study appeared to be consistent with a 'double-edged' sword view of the effects of information about MAOA × childhood abuse on sentences. The recognised increased risk of offending generally was seen as aggravating, but the causal explanation itself was seen as mitigating, reducing sentences back to around the starting point absent increased risk.

## 6.6 GENERAL DISCUSSION

These studies sought to illuminate another empirical phenomenon that does not appear to fit in with psychological theories of legal adjudication. Though previous research had suggested that the provision of causal explanations for offending behaviour actually had little or no effect on outcomes, we suspected that there might be an effect if the increased risk associated with such explanations was controlled for. In particular, we sought to test for the mitigating effect of information indicating a causal mechanism for increased risk of offending relating to the MAOA genotype × childhood abuse interaction. As predicted, participants advised that the increased risk was associated with MAOA genotype and childhood abuse consistently imposed a significantly shorter sentence. Across the experiments, of those who treated the information as relevant, there was an average decrease in sentence of 1.7 months when the causal explanation was provided (8.0 months where no causal explanation, CI = [7.6 - 8.4], and 6.3 months, CI = [6.0 - 6.7] where the risk was explained by MAOA genotype

and childhood abuse). The hypothesis that our manipulation influenced perceived blame rather than risk was supported by participants' answers to direct questions in Study 17. As predicted, the manipulation made no difference in the parole context where the task of the parole board is only to assess risk. Our research therefore provides support for the 'double-edged sword' hypothesis. It also suggests that the mitigating effects of reduced culpability may often be masked by the aggravating effects of increased risk. It is noteworthy that the causal explanation led to reduced suggested sentences even where the risk level was explicitly (quantitatively) presented as identical across conditions, and this risk was communicated as part of a forensic psychiatric report. It was the explicit identification of a genetic × environmental cause that led to reduced sentences.

As such, the research appears to confirm real-world anecdotal evidence that providing a causal explanation in terms of MAOA genotype and childhood abuse appears to reduce blameworthiness and correspondingly the seriousness of sentences. However, such information seems likely to result in shorter sentences only for those defendants who are already perceived as being of higher than risk. It therefore seems unlikely that defence lawyers would seek to adduce such information into evidence where there is otherwise no indication that their clients pose a higher risk because of the uncertain benefits.

While the findings provide *prima facie* evidence of these effects, the materials used were deliberately constrained because of the concerns about possible confounds. Replication using more naturalistic materials and with professional judges as participants would be the next logical step. If the effect can be replicated, the influence of other suspected confounds such as explicit or implicit references to psychopathy and other potentially indications of propensity to offend could be explored to test whether, as we suspect, these also moderate sentencing outcomes.

A deeper theoretical question faces judges making day-to-day decisions whether to admit such evidence and policy makers faced with developing consistent policy to address its use which is linked to the underlying question of whether this behaviour can be considered as

rational or irrational. To date, there is not a sufficiently convincing argument to settle the uncertainty one way or another. Two key philosophical camps appear to have emerged. One camp, represented by theorists such as Morse (2004, p. 180; Dershowitz, 1994; Pinker, 2009, pp. 53–54), argues that the view that causal explanations diminish blame is simply an error of reasoning. This would imply that such evidence ought not to be admitted into evidence. By contrast, other theorists such as Greene & Cohen (2004) argue that such phenomena are evidence that decision makers are switching from using common-sense psychology used for understanding agents to common-sense physics used for analysing the physical world. If the latter is correct, this could facilitate a more nuanced use of scientific research in criminal justice such that sentences are more carefully tailored to the causes of offending. Which, if either, camp is correct depends very much on our understanding of the psychology of legal decision makers, but in this specific area, our understanding remains quite rudimentary. Quite what lay and professional legal decision-makers understand when faced with causal information such as the MAOA genotype × childhood abuse interaction and how this fits into a theory of adjudication is unclear. It seems unlikely that their understanding maps precisely onto that of scientific experts researching these areas. What we know about the psychology of legal decision-makers is that causal explanations seem to have a stronger influence on sentences than verdicts and that they are very sensitive to risk. Additionally, we know that brain images and charts seem to have limited impact compared to genetic evidence. Explanations relying on the combination of genes and abuse seems to be particularly salient.

Another empirical observation that we consistently observed was the very high proportion of participants who did not deem the information about the increased risk to be relevant to sentence at all. This was consistently around half of participants in the sentencing context, compared to only around a fifth in the parole context. This seemed to be unrelated to the presence or absence of the causal information as participants appeared to object regardless of which condition they were in. Given the very stark difference between those participants who did not consider the information relevant, and those who did consider it relevant and, if given a causal explanation for it, treated it as significantly mitigating, research would be helpful to understand the nature of these two different groups.

# 7. SUMMARY AND CONCLUDING COMMENTS

We began at the outset by noting the relative dearth of descriptive psychological theories to explain legal adjudication. In particular, as many commentators note, a range of different disciplines offer prescriptive theories of adjudication, but relatively few offer descriptive theories (Baum, 1997; Hirsch, 2003, p. 602 fn16; Posner, 2008, p. 19; D. Simon, 1998, pp. 4, 32, 2010, p. 143). Adjudication also comprises a number of different, quite heterogeneous, processes which goes some way to explaining the partial focus of different theories. As noted, some focus on fact-finding, some on uncontentious disputes, some on argumentation, others on the appeal level. Adjudication also has distinctive characteristics that differentiate it from much of general psychology. One characteristic is that the context is legal, sometimes encompassing the most fundamental issues that any decision maker might be called upon to resolve. Legal decision making also encompasses the linked topics of rules and reasons. These provide a normative representation of how legal cases should be, or should have been, decided and any descriptive psychological theory needs to account for these (Braman, 2009, p. 19; Knight, 2009, p. 1538; Rowland & Carp, 1996, p. 136).

Notwithstanding these challenges, we have seen at Sections 2 and 3 that it seems possible to put together the somewhat existing disparate pieces of the puzzle into a fairly plausible theory of adjudication that accounts for much of legal adjudication. This theory meshes quite closely with common-sense views of how legal adjudication proceeds, and has characteristics of what other disciplines term 'formalist' or 'legalist' theories. The starting point is the recognition that for large swathes of the legal decision making terrain, there is a relatively high degree of consensus between adjudicators concerning what would be the correct outcome. As the leading theories, the story model, Simon's psychological model, and others correctly suggest, adjudicators make inferences from the evidence to the facts and from the facts to a decision, constrained by both black-letter law and by what others in society would find acceptable. Nonetheless, it should be noted that it may not be easy to recognise what would be an acceptable outcome. It is well recognised within law that some areas are quite troublesome to resolve because they deal with contested or evenly-balanced

questions (D. Simon, 1998, p. 19), the existing law may be unclear, and it may be hard to predict what disputes are likely to come before the courts. As such, in cases of first impression in these more challenging areas, the process of working towards the preferred approach may be an iterative process as adjudicators navigate this terrain. Here, the role of giving reasons for a decision may be characteristic of the instrumental or forward-facing function assumed by rational choice theory rooted law and economics.

We also saw how we could make a psychological theory of adjudication a little more complete by recognising the influence of values. Just as we recognise that there is a conflict between an adjudicator and an accused or litigant likely to be on the wrong end of a verdict or decision, we can also recognise that there are some topics where adjudicators' values conflict. The most obvious examples include matters of life and death such as views on the death penalty or abortion. We saw how existing psychological theories tend to focus less on situations where there are conflicts of values. Given that the inferences that adjudicators draw are otherwise oblique to observation, conflicts of values gives rise to a further role for reasons. In contentious areas where adjudicators might make decisions influenced by factors that others might find unacceptable, giving reasons may also provide a backward-looking or constraining function on adjudicators by making it more tractable for third parties to check whether the inference process set out in their reasons is internally and externally consistent.

Nonetheless, even this supplemented theory does not account for all adjudicatory behaviour. There remain areas where empirical evidence suggests that adjudicators do not seem to behave as such a theory would predict. There are occasions where adjudicators do not seem to follow the law, or where they seem to take legally impermissible influences into account. These exceptions have been the main focus of the research set out in this thesis. Current psychological theory puts many of these exceptions down to irrationality. By contrast, we have seen that some legal theory, in particular American legal realism and attitudinal theory, is more sympathetic to explanations rooted in rationality. Realist and attitudinal theories suggest that, contrary to legalist or formalist theories, legal outcomes are influenced by adjudicatory attitudes (Baum, 2006, p. 7; Bix, 2009, p. 193; Cohen, 1935, p.

840; Knight, 2009, p. 1534; Llewellyn, 1930, p. 442; Posner, 2008, pp. 19–20, 79; Robbennolt et al., 2010, p. 28; Rowland & Carp, 1996, pp. 138–139). But these legal theories still struggle to predict when attitudes will influence outcomes: is it always or only sometimes? And if only sometimes, when will this occur? And how do rules and reasons influence this? This thesis has sought to put more flesh on these bones, primarily through examining if these behaviours can be explained in more rational terms. Thus, one of the key themes running through the empirical research carried out as part of this thesis is whether the behaviours that do not fit into formalist or legalist type psychological explanations can be best explained as rational or irrational behaviours. Rational in this context means that the adjudicator is behaving in a way that is consistent with their outlook or values, rather than that that their outlook or values would be objectively acceptable to wider society. Three overarching empirical domains were examined as part of the thesis. In an approximate order of levels of certainty about the rationality of such behaviour these were: (1) the effect of legally irrelevant sympathy or antipathy on outcomes; (2) the effect of order of case presentation on outcomes; and (3) the effect of causal explanations on outcomes.

The first set of experiments in Section 4 examined the effect of sympathy or antipathy on outcomes in circumstances where these factors were legally irrelevant to the decision. The findings could seemingly be generalised to other circumstances where there is either a conflict between the values that the judge or adjudicator considers important and the applicable law or the values of others in society. Overall, the evidence appeared to be more consistent with a rational view of behaviour rather than an irrational or dual-process explanation. In particular, legally impermissible factors influenced outcomes where there was not a legally legitimate way of finding in favour of the sympathetic party and where the decision to be taken was sufficiently ambiguous that this behaviour could not be detected on an individual basis. Thus, where there were a number of issues, some of which were linked to character and some of which were not (Study 3), participants used the issues linked to character to find in favour or against a party and were apparently not influenced by character when determining the issues unrelated to character. In that study, there was no need for the participant to take the character information into account impermissibly as they could achieve

an outcome they preferred by legitimate means. However, in subsequent experiments where the issues legitimately linked to character were removed (Studies 4, 6, and 7), many participants then determined the issues that were not linked to character in accordance with character. In these studies there was no legally legitimate way to achieve the outcome participants would have been sympathetic to. The implication was that participants were only taking extra-legal factors into account where it was necessary to do so to find in favour of the side that they favoured. This suggested that the behaviour was linked to a rational sensitivity to the nature of the decision environment rather than being simply irrational behaviour.

Also contrary to the argument that the behaviour was caused by the complexity of the task, was the observation the behaviour manifested itself even in the very simplest of tasks where there were only one or two issues to be determined (Studies 1 and 2). Furthermore, the observation that these extralegal influences did not occur in more complicated studies with as many as 6 issues to be determined (Study 3) suggested that complexity was not the most influential factor.

Participants also appeared to be quite sophisticated in how they used this collateral information. In a series of experiments, information relating to character was provided at an interim stage and a final stage. When participants were asked to assess the issues at an interim stage, their responses were influenced by character (Study 7). When participants were asked to assess the issues at a final stage after a dual (opposing) character manipulation, their responses were in accordance with the manipulation seen second (Studies 6 and 7). If participants were asked to give reasons for their assessment, this effectively eliminated the impermissible effect of character at the preliminary indication stage, but not the final decision stage (Study 6). This suggested some sensitivity to the risk of the impermissible use of this information being highlighted before all the evidence was complete. Relatedly, where participants gave preliminary assessments that were influenced by the first character manipulation, they stuck by these assessments for the final decision after the second character manipulation (Study 7), even where this meant that they were effectively punishing the more sympathetic party. Participants therefore appeared very alive to whether their reliance on

legally impermissible factors could be detected, and altered their behaviour to prevent this, even if it ultimately led to a result which they would not have wanted. As such, this suggested a much more rational and sophisticated picture of adjudicatory behaviour than irrationality caused by complexity and lack of cognitive capacity.

The second set of experiments in Section 5, examining the effect of order of case presentation on case outcomes, linked to some of the themes examined in Section 4. More specifically, there was a suggestion in the first set of experiments that once participants had committed to a view that was impermissibly influenced by character, they subsequently stuck by that view even if it ended up disadvantaging the sympathetic party (Study 7). The implication was that participants did not want to appear inconsistent as that would indicate that they had taken impermissible factors into account. This is the same rationale as many commentators use to explain the order effects seen in paired moral psychology dilemmas: the suggestion is that decision makers want to appear consistent in their decisions because inconsistency suggests bias, carelessness, or ignorance (Engel, 2006, p. 250). Our second set of experiments sought to extend the findings from moral psychology research to the legal domain. Numerous replicated experiments have shown that there is an order effect with the presentation of moral dilemmas to participants such that it matters whether the dilemmas are presented in the order A>B or B>A where dilemma A is generally approved of in isolation and dilemma B is generally disapproved of in isolation. We sought to replicate these findings in the legal context by presenting similar paired legal cases in different orders. As predicted, we were able to demonstrate order effects with both civil law (Study 8) and criminal law cases (Studies 9, 10, and 11). We also incidentally replicated the asymmetrical patterns seen in moral psychology research whereby one of the dilemmas is stabile, uninfluenced by order of presentation, whereas the other is labile, responses differing depending on whether it is presented first or last (Studies 8, 9, 10, and 11).

Notably, the direction of the order effects that we found were different to previous moral psychology research. Whereas previous research had demonstrated that responses to the labile dilemma presented last tend to move closer to responses to the dilemma that

preceded it, we found that in at least some instances, responses to the labile dilemma moved further away from responses to the dilemma that preceded it (Studies 9, 10, and 11). Put differently, participant responses appeared to become less consistent rather than more consistent. Though it seems likely that there are circumstances where consistency is a relevant factor affecting responses (for example, Study 7), here it seemed that consistency was not the most influential consideration for participants in Studies 9 to 13.

These empirical findings also seem to be in conflict with theories that suggest that order effects are due to labile scenarios having a more ambiguous underlying causal structure. While causal structure may be relevant to 'trolley' type experiments due to the obvious importance of the underlying causal structure in those dilemmas, they seem less relevant to our shipwreck type experiments. Here the key factors appeared to be the identity of the perpetrators and victim and the selection procedure adopted. Nonetheless, theories that assume the importance of the underlying causal structure would seem to imply that responses to labile or ambiguous scenarios would be influenced so as to be more like the stabile or unambiguous scenarios that preceded them. This was not what we found, suggesting that other explanations should be sought.

The remaining category of explanation that appeared relevant was salience type explanations that assume that the case presented first highlights or makes salient factors that the participant did not previously take sufficiently into account at the outset. Given that such information could cast the labile scenario in either a more favourable or less favourable light, it seemed possible that this could explain responses to labile cases sometimes becoming more dissimilar from the stabile cases that preceded them. However, if salience is the best explanation, the further studies that we carried out provided limited support for this. Drawing participants' attention to the factors presumed to be salient by disclosing the possibility of a fairer method of selection (Study 11) did not apparently affect responses as salience explanations would predict. Equally, on the basis of previous research that suggested that groups are much more likely to identify salient factors than individual decision makers, we tested responses of participants able to deliberate as a group against those required to decide

in isolation (Study 12). If the order effects seen in labile cases are due to participants initially failing to appreciate the possibility of a fairer method of selection, we would have expected groups to be more likely to identify this and therefore to assess the cases where a fair method of selection was not adopted more harshly. But responses in the group condition were practically identical to the individual condition. Finally, we examined whether a condition with a lively illustration of a scrupulously fair means of selection taken from a real life shipwreck would also make participants assess a case where an unfair means of selection was adopted more harshly (Study 13). However, notwithstanding that participants in this condition were much more likely, when asked, to identify a fairer means of selection, this apparently did not make them judge the perpetrators more harshly.

Overall, the second set of experiments replicated and extended the order effects seen in the moral decision-making context to the legal context, indicating that order effects in both directions seem likely to occur in both civil and criminal contexts. In addition, the theories propounded so far to explain these order effects do not seem to be particularly compatible with the empirical results we have found, suggesting that other explanations may be required to account for these order effects in both the moral and legal spheres. Thus it is currently too early to discern whether these behaviour are evidence of rational or irrational behaviour.

The third set of experiments in Section 6 concerned the effect of causal information on legal outcomes, and was specifically focussed on addressing an empirical question. This question was whether providing causal information about an offender's MAOA genotype and childhood abuse had mitigating effects on sentence. Many researchers had concluded that this type of information had little or no effect on sentence, but we wished to examine whether any mitigating effects were being cancelled out by the aggravating effects linked to the increased risk posed by individuals with these characteristics. To do this we considered two criminal justice contexts, parole and sentencing.

In the parole context, a firm majority of around four-fifths of participants consistently considered the information about the offender relevant to their risk assessment (Studies 14

and 17). Perhaps unsurprisingly, all participants considered the information to increase the risk posed by the accused, but there was no significant difference between whether or not they were provided with causal information to explain that increased risk (Studies 14 and 17).

By contrast, there was a very different picture in the sentencing context. Around half of participants did not think the risk information was relevant to sentencing (Studies 14, 15, 16, and 17), though this did not seem to be associated with objections to the use of genetic information as there was no significant difference in relevance assessments depending on whether or not participants were provided with the causal information explaining the increased risk. Participants given only information about the increased risk posed by the accused invariably gave much longer sentences than the baseline sentence otherwise indicated as appropriate. Participants also given the causal information about MAOA genotype and childhood abuse to explain this increased risk gave a much shorter sentence than those simply provided with the increased risk. What was interesting was that the mean sentence imposed by those given the causal information was very close to the initial baseline sentence indicated at the outset because this might imply that the two phenomena, risk and causal explanation, may often come quite close to balancing each other out in experimental or real-world contexts as suggested by the 'double-edged sword' hypothesis.

Our studies therefore imply that the anecdotal and survey evidence of the use of this type of causal information by real-world defence lawyers may be justified. While the increased risk posed by individuals with these characteristics may be associated with aggravating effects, these consequences seem to be effectively counterbalanced by the mitigating effects. Given that such individuals may well be considered an elevated risk based on other information, it seems probable that the reliance on this information may generally be advantageous, on balance, to the defence.

Determining whether these effects of causal information on outcomes could be considered rational or irrational behaviours is fairly challenging, given that it raises quite fundamental philosophical questions relating to issues such as free will. One leading view is

that they are irrational behaviours where mechanisms for fact-finding and decision-making go awry, caused by decision makers assessing questions relating to people and agents using cognitive mechanisms adapted for assessing objects and events. Against this, not every theorist accepts that human cognitive systems are so modular or informationally encapsulated (Fodor, 1987, p. 139; Okasha, 2002, p. 168). Given the considerable theoretical work required to advance an answer to this question, it would be premature to try to categorise these effects as rational or irrational.

Overall, we have seen how it is possible to advance the psychological theory of adjudication on the basis of existing theory and research to make it more comprehensive; to extend this theory to encompass some of the effects of extra-legal values or information that have previously been attributed to irrationality; and to identify a number of robust empirical findings linked to order of case presentation and causal information that are not well explained by current psychological theory.

# REFERENCES


Airedale NHS Trust v Bland, [1993] AC 789 (HL 1993).

Allard, J., & Fortin, M.-C. (2017). Organ Donation After Medical Assistance in Dying or Cessation of Life-Sustaining Treatment Requested by Conscious Patients: The Canadian Context. *Journal of Medical Ethics*, *43*(9), 601–605. https://doi.org/10.1136/medethics-2016-103460

Allen, C. H., Vold, K., Felsen, G., Blumenthal-Barby, J. S., & Aharoni, E. (2019). Reconciling the Opposing Effects of Neurobiological Evidence on Criminal Sentencing Judgments. *PLOS ONE*, *14*(1), e0210584. https://doi.org/10.1371/journal.pone.0210584

Allen, R. J. (2000). Common Sense, Rationality, and the Legal Process. *Cardozo Law Review*, *22*, 1417.

Anderson, N. H. (1959). Test of a Model for Opinion Change. *The Journal of Abnormal and Social Psychology*, *59*(3), 371.

Anderson, T., Schum, D., & Twining, W. (2005). *Analysis of Evidence* (2nd ed.). Cambridge University Press. http://ebooks.cambridge.org/ref/id/CBO9780511610585

Anonymous. (2013, August 9). Student Asks to Donate Organs Before Committing Suicide. *The Daily Telegraph*. https://www.telegraph.co.uk/news/10233221/Student-asks-to-donate-organs-before-committing-suicide.html

Aono, D., Yaffe, G., & Kober, H. (2019). Neuroscientific Evidence in the Courtroom: A Review. *Cognitive Research: Principles and Implications*, *4*(1), 40. https://doi.org/10.1186/s41235-019-0179-y

Appelbaum, P. S., & Scurich, N. (2014). Impact of Behavioral Genetic Evidence on the Adjudication of Criminal Behavior. *Journal of the American Academy of Psychiatry and the Law Online*, *42*(1), 91–100.

Appelbaum, P. S., Scurich, N., & Raad, R. (2015). Effects of Behavioral Genetic Evidence on Perceptions of Criminal Responsibility and Appropriate Punishment. *Psychology, Public Policy, & Law*, *21*(2), 134–144. https://doi.org/10.1037/law0000039

Appiah, A. (2008). *Experiments in Ethics*. Harvard University Press.

Aspinwall, L. G., Brown, T. R., & Tabery, J. (2012). The Double-Edged Sword: Does Biomechanism Increase or Decrease Judges' Sentencing of Psychopaths? *Science*, *337*(6096), 846–849. https://doi.org/10.1126/science.1219569

Austin, J. L. (1956). A Plea for Excuses: The Presidential Address. *Proceedings of the Aristotelian Society*, *57*, 1–30.

Ayer, A. J. (1946). Freedom and Necessity. In *Philosophical Essays*. Palgrave MacMillan.

Bahník, Š., Mussweiler, T., & Strack, F. (2022). Anchoring Effect. In R. F. Pohl (Ed.), *Cognitive Illusions: Intriguing Phenomena in Judgement, Thinking and Memory*. Routledge. https://doi.org/10.4324/9781003154730-16

Baker, C. L., Saxe, R. R., & Tenenbaum, J. B. (2011). Bayesian Theory of Mind: Modeling Joint Belief-Desire Attribution. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2469–2474). Cognitive Science Society.

Baker, D. A., Schweitzer, N. J., Risko, E. F., & Ware, J. M. (2013). Visual Attention and the Neuroimage Bias. *PLOS ONE*, *8*(9), e74449. https://doi.org/10.1371/journal.pone.0074449

Baker, J. H. (2002). *An Introduction to English Legal History* (4th edition). Butterworths.

Baron, J. (2008). *Thinking and Deciding* (4th ed.). Cambridge University Press.

Baum, L. (1997). *The Puzzle of Judicial Behavior*. University of Michigan Press. https://muse.jhu.edu/book/6316

Baum, L. (2006). *Judges and Their Audiences: A Perspective on Judicial Behavior*. Princeton University Press.

Bayes, T., & Price, R. (1763). An Essay towards Solving a Problem in the Doctrine of Chances. *Philosophical Transactions*, *53*, 370–418. https://doi.org/10.1098/rstl.1763.0053

Beale, J. H. (1916). *A Treatise on the Conflict of Laws or, Private International Law*. Cambridge [Mass.] : Harvard University Press. http://archive.org/details/cu31924022034684

Becker, G. S. (1993). Nobel Lecture: The Economic Way of Looking at Behavior. *Journal of Political Economy*, *101*(3), 385–409. https://doi.org/10.2307/2138769

Berryessa, C. M. (2016). Judges' Views on Evidence of Genetic Contributions to Mental Disorders in Court. *The Journal of Forensic Psychiatry & Psychology*, *27*(4), 586–600. https://doi.org/10.1080/14789949.2016.1173718

Berryessa, C. M., Coppola, F., & Salvato, G. (2021). The Potential Effect of Neurobiological Evidence on the Adjudication of Criminal Responsibility of Psychopathic Defendants in Involuntary Manslaughter Cases. *Psychology, Crime & Law*, *27*(2), 140–158. https://doi.org/10.1080/1068316X.2020.1780590

Bex, F. J. (2011). *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory*. Springer Science & Business Media.

Binmore, K. G. (2011). *Rational Decisions*. Princeton University Press.

Bix, B. H. (2009). *Jurisprudence: Theory and Context* (5th edition). Sweet & Maxwell.

Blakey, R., & Kremsmayer, T. P. (2018). Unable or Unwilling to Exercise Self-control? The Impact of Neuroscience on Perceptions of Impulsive Offenders. *Frontiers in Psychology*, *8*, 2189. https://doi.org/10.3389/fpsyg.2017.02189

Blume, J. H., Garvey, S. P., & Johnson, S. L. (2000). Future Dangerousness in Capital Cases: Always at Issue. *Cornell Law Review*, *86*(2), 397–410.

Boden, R. (Director). (1989, October 5). Plan B: Corporal Punishment (Episode 2). In *Blackadder Goes Forth*. BBC.

Boiney, L. G., Kennedy, J., & Nye, P. (1997). Instrumental Bias in Motivated Reasoning: More When More Is Needed. *Organizational Behavior and Human Decision Processes*, *72*(1), 1–24.

Bollen, J., Jongh, W. de, Hagenaars, J., Dijk, G. van, Hoopen, R. ten, Ysebaert, D., Ijzermans, J., Heurn, E. van, & Mook, W. van. (2016). Organ Donation After Euthanasia: A Dutch Practical Manual. *American Journal of Transplantation*, *16*(7), 1967–1972. https://doi.org/10.1111/ajt.13746

Bollen, J., Smaalen, T. van, Hoopen, R. ten, Heurn, E. van, Ysebaert, D., & Mook, W. van. (2017). Potential Number of Organ Donors After Euthanasia in Belgium. *JAMA*, *317*(14), 1476–1477. https://doi.org/10.1001/jama.2017.0729

Bordalo, P., Gennaioli, N., & Shleifer, A. (2015). Salience Theory of Judicial

Decisions. *The Journal of Legal Studies*, *44*(S1), S7–S33. https://doi.org/10.1086/676007

Bovens, L., & Hartmann, S. (2004). *Bayesian Epistemology*. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/0199269750.001.0001/acprof-9780199269754

Braman, E. (2006). Reasoning on the Threshold: Testing the Separability of Preferences in Legal Decision Making. *The Journal of Politics*, *68*(2), 308–321. JSTOR. https://doi.org/10.1111/j.1468-2508.2006.00408.x

Braman, E. (2009). *Law, Politics, and Perception: How Policy Preferences Influence Legal Reasoning*. University of Virginia Press. http://muse.jhu.edu/book/5263

Braman, E., & Nelson, T. E. (2007). Mechanism of Motivated Reasoning? Analogical Perception in Discrimination Disputes. *American Journal of Political Science*, *51*(4), 940–956.

Brenner, S., & Spaeth, H. J. (1995). *Stare Indecisis: The Alteration of Precedent on the Supreme Court, 1946-1992*. Cambridge University Press.

Brigham, J. (1978). *Constitutional Language: An Interpretation of Judicial Decision: 17*. Praeger.

Broeder, D. W. (1959). The University of Chicago Jury Project Special Feature on Damages. *Nebraska Law Review*, *38*(3), 744–760.

Bruner, J. S. (1973). *Beyond the Information Given: Studies in the Psychology of Knowing*. George Allen & Unwin.

Byrd, A. L., & Manuck, S. B. (2014). MAOA, Childhood Maltreatment, and Antisocial Behavior: Meta-analysis of a Gene-Environment Interaction. *Biological Psychiatry*, *75*(1), 9–17. https://doi.org/10.1016/j.biopsych.2013.05.004

Caspi, A., McCLay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., Taylor, A., & Poulton, R. (2002). Role of Genotype in the Cycle of Violence in Maltreated Children. *Science*, *297*(5582), 851–854.

Catley, P., & Claydon, L. (2015). The Use of Neuroscientific Evidence in the Courtroom by Those Accused of Criminal Offenses in England and Whales. *Journal of Law and the Biosciences*, *2*(3), 510–549.

Chapman, G. B., & Bornstein, B. H. (1996). The More You Ask for, the More You

Get: Anchoring in Personal Injury Verdicts. *Applied Cognitive Psychology*, *10*(6), 519–540. https://doi.org/10.1002/(SICI)1099-0720(199612)10:6<519::AID-ACP417>3.0.CO;2-5

Chapman, G. B., & Johnson, E. J. (1999). Anchoring, Activation, and the Construction of Values. *Organizational Behavior and Human Decision Processes*, *79*(2), 115–153. https://doi.org/10.1006/obhd.1999.2841

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic Models of Cognition: Conceptual Foundations. *Trends in Cognitive Sciences*, *10*(7), 287–291. https://doi.org/10.1016/j.tics.2006.05.007

Cheung, B. Y., & Heine, S. J. (2015). The Double-Edged Sword of Genetic Accounts of Criminality: Causal Attributions From Genetic Ascriptions Affect Legal Decision Making. *Personality and Social Psychology Bulletin*, *41*(12), 1723–1738. https://doi.org/10.1177/0146167215610520

Chomsky, N. (1957). *Syntactic Structures* (2nd ed.). Mouton.

Churchland, P. M. (1981). Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, *78*(2), 67–90. https://doi.org/10.2307/2025900

Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. MIT Press.

Cohen, F. S. (1935). Transcendental Nonsense and the Functional Approach. *Columbia Law Review*, *35*(6), 809–849. https://doi.org/10.2307/1116300

Coleman, J. L., & Leiter, B. (1993). Determinacy, Objectivity, and Authority. *University of Pennsylvania Law Review*, *142*(2), 549–637. https://doi.org/10.2307/3312546

Confer, J. A., & Chopik, W. J. (2019). Behavioral Explanations Reduce Retributive Punishment but Not Reward: The Mediating Role of Conscious Will. *Consciousness and Cognition*, *75*, 102808.

Cooter, R., & Ulen, T. S. (2012). *Law & Economics* (6th ed.). Addison-Wesley.

Costabile, K. A., & Klein, S. B. (2005). Finishing Strong: Recency Effects in Juror Judgments. *Basic and Applied Social Psychology*, *27*(1), 47–58. https://doi.org/10.1207/s15324834basp2701_5

Cox, A., & Miles, T. J. (2008). Judging the Voting Rights Act. *Columbia Law Review*, *108*(1), 1–54.

Cushman, F., Young, L., & Hauser, M. (2006). The Role of Conscious Reasoning and Intuition in Moral Judgment. *Psychological Science*, *17*(12), 1082.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness. *Economic Theory*, *33*(1), 67–80. https://doi.org/10.1007/s00199-006-0153-z

Dawid, A. P., & Evett, I. W. (1997). Using a Graphical Method to Assist the Evaluation of Complicated Patterns of Evidence. *Journal of Forensic Sciences*, *42*(2), 226–231. https://doi.org/10.1520/JFS14102J

de Kogel, C. H., & Westgeest, E. J. M. C. (2015). Neuroscientific and Behavioral Genetic Information in Criminal Cases in the Netherlands. *Journal of Law and the Biosciences*, *2*(3), 580–605. https://doi.org/10.1093/jlb/lsv024

Dennett, D. C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. Clarendon Press ; Oxford University Press.

Dennett, D. C. (1987). Cognitive Wheels: The Frame Problem of AI. In Z. W. Pylyshyn (Ed.), *The Robot's Dilemma: Frame Problem in Artificial Intelligence* (pp. 41–64). Ablex Publishing Corporation.

Denno, D. W. (2015). The Myth of the Double-Edged Sword: An Empirical Study of Neuroscience Evidence in Criminal Cases. *Boston College Law Review*, *56*(2), 493–552.

Dershowitz, A. M. (1994). *The Abuse Excuse: And Other Cop-Outs, Sob Stories, and Evasions of Responsibility*. Back Bay Books.

Devine, D. J. (2012). *Jury Decision Making: The State of the Science*. NYU Press.

Dhami, M. K. (2003). Psychological Models of Professional Decision Making. *Psychological Science*, *14*(2), 175–180. https://doi.org/10.1111/1467-9280.01438

Dhami, M. K., & Thomson, M. E. (2012). On the Relevance of Cognitive Continuum Theory and Quasirationality for Understanding Management Judgment and Decision Making. *European Management Journal*, *30*(4), 316–326. https://doi.org/10.1016/j.emj.2012.02.002

Diamond, S. S., & Rose, M. R. (2018). The Contemporary American Jury. *Annual Review of Law and Social Science*, *14*(1), 239–258. https://doi.org/10.1146/annurev-lawsocsci-110316-113618

Dijk, G. van, Bruchem-Visser, R. van, & Beaufort, I. de. (2018). Organ Donation

After Euthanasia, Morally Acceptable Under Strict Procedural Safeguards. *Clinical Transplantation*, *32*(8), e13294. https://doi.org/10.1111/ctr.13294

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Chapter 10 Motivated Moral Reasoning. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 50, pp. 307–338). Academic Press. http://www.sciencedirect.com/science/article/pii/S0079742108004106

Duhem, P. M. M. (1914). *The Aim and Structure of Physical Theory* (P. P. Wiener, Trans.). Princeton UP; Oxford UP.

Duncker, K. (1945). *On Problem-Solving* (J. F. Dashiell, Ed.; L. S. Lees, Trans.). American Psychological Association.

Dworkin, R. (1977). *Taking Rights Seriously*. Harvard University Press.

Dworkin, R. (1986). *Law's Empire*. Belknap Press.

Edwards, W. (1991). Influence Diagrams, Bayesian Imperialism, and the Collins Case: An Appeal to Reason Decision and Interference Litigation. *Cardozo Law Review*, *13*(Issues 2-3), 1025–1074.

Eisenberg, M. A. (1978). Participation, Responsiveness, and the Consultative Process: An Essay for Lon Fuller. *Harvard Law Review*, *92*(2), 410–432.

Eisenberg, T., & Hans, V. P. (2008). Taking a Stand on Taking the Stand: The Effect of a Prior Criminal Record on the Decision to Testify and on Trial Outcomes. *Cornell Law Review*, *94*(6), 1353–1390.

Ellickson, R. C. (1986). Of Coase and Cattle: Dispute Resolution among Neighbors in Shasta County. *Stanford Law Review*, *38*(3), 623–687. https://doi.org/10.2307/1228561

Ellsworth, P. C. (1989). Are Twelve Heads Better than One? *Law and Contemporary Problems*, *52*(4), 205–224. https://doi.org/10.2307/1191911

Engel, C. (2006). Inconsistency in the Law: In Search of a Balanced Norm. In C. Engel & L. Daston (Eds.), *Is There Value in Inconsistency?* (pp. 223–281). Nomos Verlagsgesellschaft.

Engel, C. (2008). Learning the Law. *Journal of Institutional Economics*, *4*(03), 275. https://doi.org/10.1017/S1744137408001094

Engel, C., Timme, S., & Glöckner, A. (2020). Coherence-Based Reasoning and Order Effects in Legal Judgments. *Psychology, Public Policy, and Law*, *26*(3), 333–352.

https://doi.org/10.1037/law0000257

Englich, B., & Mussweiler, T. (2001). Sentencing Under Uncertainty: Anchoring Effects in the Courtroom. *Journal of Applied Social Psychology*, *31*(7), 1535–1551. https://doi.org/10.1111/j.1559-1816.2001.tb02687.x

Englich, B., Mussweiler, T., & Strack, F. (2005). The Last Word in Court—A Hidden Disadvantage for the Defense. *Law and Human Behavior*, *29*(6), 705–722. https://doi.org/10.1007/s10979-005-8380-7

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. https://doi.org/10.1177/0146167205282152

Epley, N. (2004). A Tale of Tuned Decks? Anchoring as Accessibility and Anchoring as Adjustment. In *Blackwell Handbook of Judgment and Decision Making* (pp. 240–257). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470752937.ch12

Epley, N., & Gilovich, T. (2001). Putting Adjustment Back in the Anchoring and Adjustment Heuristic: Differential Processing of Self-Generated and Experimenter-Provided Anchors. *Psychological Science (0956-7976)*, *12*(5), 391. https://doi.org/10.1111/1467-9280.00372

Epstein, L., Landes, W. M., & Posner, R. A. (2013). *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice*. Cambridge, Mass: Harvard University Press. https://www.jstor.org/stable/10.2307/j.ctt2jbs80

Farahany, N. A. (2015). Neuroscience and Behavioral Genetics in Us Criminal Law: An Empirical Analysis. *Journal of Law and the Biosciences*, *2*(3), 485–509. https://doi.org/10.1093/jlb/lsv059

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behavior Research Methods*, *39*(2), 175–191.

Feldman, Y., Schurr, A., & Teichman, D. (2016). Anchoring Legal Standards. *Journal of Empirical Legal Studies*, *13*. https://doi.org/10.1111/jels.12116

Fenton, N. E., & Neil, M. (2013). *Risk Assessment and Decision Analysis with*

*Bayesian Networks*. CRC Press.

Feresin, E. (2009). Lighter Sentence for Murderer with 'Bad Genes'. *Nature News*. https://doi.org/10.1038/news.2009.1050

Feresin, E., & Owens, B. (2011, September 1). *Nature News Blog: Italian Court Reduces Murder Sentence Based on Neuroimaging Data*. http://blogs.nature.com/news/2011/09/italian_court_reduces_murder_s.html

Festinger, L. (1962). Cognitive Dissonance. *Scientific American*, *207*(4), 93. https://doi.org/10.1038/scientificamerican1062-93

Feteris, E. T. (2017). *Fundamentals of Legal Argumentation: A Survey of Theories on the Justification of Judicial Decisions* (2nd ed, Vol. 1). Springer Netherlands. https://doi.org/10.1007/978-94-024-1129-4

Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. MIT Press.

Fodor, J. A. (1987). Modules, Frames, Fridgeons, Sleeping Dogs, and the Music of the Spheres. In Z. W. Pylyshyn (Ed.), *The Robot's Dilemma: Frame Problem in Artificial Intelligence* (pp. 139–150). Ablex Publishing Corporation.

Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, *5*, 5–15.

Fried, C. B. (1996). Bad Rap for Rap: Bias in Reactions to Music Lyrics. *Journal of Applied Social Psychology*, *26*(23), 2135–2146. https://doi.org/10.1111/j.1559-1816.1996.tb01791.x

Fuller, L. L. (1957). Positivism and Fidelity to Law—A Reply to Professor Hart. *Harvard Law Review*, *71*(4), 630–672.

Fuller, L. L. (1978). The Forms and Limits of Adjudication. *Harvard Law Review*, *92*(2), 353–409.

Furgeson, J. R., Babcock, L., & Shane, P. M. (2008). Do a Law's Policy Implications Affect Beliefs about Its Constitutionality? An Experimental Test. *Law and Human Behavior*, *32*(3), 219–227.

Furnham, A. (1986). The Robustness of the Recency Effect: Studies Using Legal Evidence. *Journal of General Psychology*, *113*(4), 351.

https://doi.org/10.1080/00221309.1986.9711045

Fuss, J., Dressing, H., & Briken, P. (2015). Neurogenetic Evidence in the Courtroom: A Randomised Controlled Trial with German Judges. *Journal of Medical Genetics*, *52*(11), 730–737. https://doi.org/10.1136/jmedgenet-2015-103284

George, T. E., & Epstein, L. (1992). On the Nature of Supreme Court Decision Making. *The American Political Science Review*, *86*(2), 323–337. JSTOR. https://doi.org/10.2307/1964223

Gigerenzer, G. (2000). *Adaptive Thinking: Rationality in the Real World*. Oxford University Press.

Gigerenzer, G., & Engel, C. (2006). Law and Heuristics: An Interdisciplinary Venture. In G. Gigerenzer & C. Engel (Eds.), *Heuristics And the Law*. MIT Press.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gilbo, N., Jochmans, I., Jacobs-Tulleneers-Thevissen, D., Wolthuis, A., Sainz-Barriga, M., Pirenne, J., & Monbaliu, D. (2019). Survival of Patients With Liver Transplants Donated After Euthanasia, Circulatory Death, or Brain Death at a Single Center in Belgium. *JAMA*, *322*(1), 78–80. https://doi.org/10.1001/jama.2019.6553

Glöckner, A., & Engel, C. (2013). Can We Trust Intuitive Jurors? Standards of Proof and the Probative Value of Evidence in Coherence-Based Reasoning. *Journal of Empirical Legal Studies*, *10*(2), 230–252. https://doi.org/10.1111/jels.12009

Glymour, C. (1992). Invasion of the Mind Snatchers. In R. N. Giere (Ed.), *Cognitive Models of Science* (pp. 465–474). U of Minnesota Press.

Goff, R. (1999). Appendix: The Search for Principle. In W. Swadling & G. H. Jones (Eds.), *The Search for Principle: Essays in Honour of Lord Goff of Chieveley* (pp. 313–329). Oxford University Press.

Gordon, N., & Greene, E. (2018). Nature, Nurture, and Capital Punishment: How Evidence of a Genetic–Environment Interaction, Future Dangerousness, and Deliberation Affect Sentencing Decisions. *Behavioral Sciences & the Law*, *36*(1), 65–83. https://doi.org/10.1002/bsl.2306

Greene, E., & Cahill, B. S. (2012). Effects of Neuroimaging Evidence on Mock Juror

Decision Making Special Issue: Current Divisions. *Behavioral Sciences & the Law*, *30*(3), 280–296.

Greene, E., Chopra, S. R., Kovera, M. B., Penrod, S. D., Rose, V. G., Schuller, R., & Studebaker, C. A. (2006). Jurors and Juries: A Review of the Field. In J. R. P. Ogloff (Ed.), *Taking Psychology and Law into the Twenty-First Century* (pp. 225–285). Springer Science & Business Media.

Greene, J. D., & Cohen, J. D. (2004). For the Law, Neuroscience Changes Nothing and Everything. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1451), 1775–1785. https://doi.org/10.1098/rstb.2004.1546

Grice, P. (1991). Postwar Oxford Philosophy. In *Studies in the Way of Words* (New Ed edition, pp. 171–180). Harvard University Press.

Guillen Gonzalez, D., Bittlinger, M., Erk, S., & Müller, S. (2019). Neuroscientific and Genetic Evidence in Criminal Cases: A Double-Edged Sword in Germany but Not in the United States? *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02343

Gurley, J. R., & Marcus, D. K. (2008). The Effects of Neuroimaging and Brain Injury on Insanity Defenses. *Behavioral Sciences & the Law*, *26*(1), 85–98.

Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2001). Inside the Judicial Mind. *Cornell Law Review*, *86*(4), 777–777.

Hahn, U., Harris, A. J. L., & Corner, A. (2015). Public Reception of Climate Science: Coherence, Reliability, and Independence. *Topics in Cognitive Science*, n/a-n/a. https://doi.org/10.1111/tops.12173

Haidt, J. (2001). The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review*, *108*(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Hart, H. L. A. (1958). Positivism and the Separation of Law and Morals. *Harvard Law Review*, *71*(4), 593–629. https://doi.org/10.2307/1338225

Hart, H. L. A. (1961). *The Concept of Law* (2nd ed. / with a postscript edited by Penelope A. Bulloch and Joseph Raz.). Clarendon Press.

Hart, H. L. A., & Honoré, T. (1985). *Causation in the Law*. Oxford University Press.

Hastie, R., Penrod, S., & Pennington, N. (1983). *Inside the Jury*. Harvard University

Press.

Hastie, R., Schkade, D. A., & Payne, J. W. (1999). Juror Judgments in Civil Cases: Effects of Plaintiff's Requests and Plaintiff's Identity on Punitive Damage Awards. *Law and Human Behavior*, *23*(4), 445–470.

Heiner, R. A. (1986). Imperfect Decisions and the Law: On the Evolution of Legal Precedent and Rules. *The Journal of Legal Studies*, *15*(2), 227–261.

Heller, T. C. (1979). Is the Charitable Exemption from Property Taxation an Easy Case? General Concern About Economics. In D. L. Rubinfeld (Ed.), *Essays on the Law and Economics of Local Governments* (pp. 183-???). Urban Institute.

Hempel, C. G. (Carl G. (1966). *Philosophy of Natural Science*. Prentice-Hall.

Hertwig, R. (2006). Do Legal Rules Rule Behavior? In *Heuristics and the Law* (pp. 391–410). Dahlem University Press.

Higgins, R. S., & Rubin, P. H. (1980). Judicial Discretion. *Journal of Legal Studies*, *9*, 129.

Hirsch, A. J. (2003). Cognitive Jurisprudence. *Southern California Law Review*, *76*, 1331.

Ho, H. L. (2008). *A Philosophy of Evidence Law*. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199228300.001.0001/acprof-9780199228300

Holmes, O. (1897). The Path of the Law. *Harvard Law Review*, *10*(8), 457–478. https://doi.org/10.2307/1322028

Holyoak, K. J., & Simon, D. (1999). Bidirectional Reasoning in Decision Making by Constraint Satisfaction. *Journal of Experimental Psychology: General*, *128*(1), 3–31. https://doi.org/10.1037/0096-3445.128.1.3

Horne, Z., & Livengood, J. (2017). Ordering Effects, Updating Effects, and the Specter of Global Skepticism. *Synthese; Dordrecht*, *194*(4), 1189–1218. http://dx.doi.org.libproxy.ucl.ac.uk/10.1007/s11229-015-0985-9

Horst, S. (2011). The Computational Theory of Mind. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2011). http://plato.stanford.edu/archives/spr2011/entries/computational-mind/

Hsee, C. K. (1996). Elastic Justification: How Unjustifiable Factors Influence Judgments. *Organizational Behavior and Human Decision Processes*, *66*(1), 122–129. https://doi.org/10.1006/obhd.1996.0043

Jeremy Bentham. (1780). *An Introduction to the Principles of Morals and Legislation*. Clarendon Press, Clarendon. http://dx.doi.org/10.1093/actrade/9780198205166.book.1

Johnson, C. A. (1987). Law, Politics, and Judicial Decision Making: Lower Federal Court uses of Supreme Court Decisions. *Law and Society Review*, *21*(2), 325–340.

Jones, E. E., & Davis, K. E. (1965). From Acts To Dispositions The Attribution Process In Person Perception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 2, pp. 219–266). Academic Press. https://doi.org/10.1016/S0065-2601(08)60107-0

Jones, M., & Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the Explanatory Status and Theoretical Contributions of Bayesian Models of Cognition. *Behavioral and Brain Sciences*, *34*(04), 169–188. https://doi.org/10.1017/S0140525X10003134

Kadane, J. B., & Schum, D. A. (1996). *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. Wiley.

Kahan, D. M., Hoffman, D. A., Braman, D., & Evans, D. (2012). They Saw a Protest: Cognitive Illiberalism and the Speech-Conduct Distinction. *Stanford Law Review*, *64*(4), 851–906.

Kahneman, D. (2012). *Thinking, Fast and Slow*. Penguin.

Karni, E. (2005). State-Dependent Preferences. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *A Dictionary of Economic Theory and Doctrine*. Palgrave MacMillan.

Kassin, S. M., & Sommers, S. R. (1997). Inadmissible Testimony, Instructions to Disregard, and the Jury: Substantive Versus Procedural Considerations. *Personality and Social Psychology Bulletin*, *23*(10), 1046–1054. https://doi.org/10.1177/01461672972310005

Katz, C. M., & Spohn, C. C. (1995). The Effect of Race and Gender on Bail Outcomes: A Test of an Interactive Model. *American Journal of Criminal Justice*, *19*(2), 161–184. https://doi.org/10.1007/BF02885913

Kelley, H. H. (1973). The Processes of Causal Attribution. *American Psychologist*,

*28*(2), 107–128.

Kelley, H. H. (1987). Attribution in Social Interaction. In E. E. Jones, D. E. Kanhouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the Causes of Behavior*. Lawrence Erlbaum Associates, Publishers.

Kelley, S. E., Edens, J. F., Mowle, E. N., Penson, B. N., & Rulseh, A. (2019). Dangerous, Depraved, and Death-Worthy: A Meta-Analysis of the Correlates of Perceived Psychopathy in Jury Simulation Studies. *Journal of Clinical Psychology*, *75*(4), 627–643. https://doi.org/10.1002/jclp.22726

Kelman, M. (2011). *The Heuristics Debate*. Oxford University Press.

Kelman, M., Rottenstreich, Y., & Tversky, A. (1996). Context-Dependence in Legal Decision Making. *Journal of Legal Studies*, *25*, 287–318.

Kerstholt, J. H., & Jackson, J. L. (1998). Judicial Decision Making: Order of Evidence Presentation and Availability of Background Information. *Applied Cognitive Psychology*, *12*(5), 445–454. https://doi.org/10.1002/(SICI)1099-0720(199810)12:5<445::AID-ACP518>3.0.CO;2-8

Kim, J., Boytos, A., Seong, Y., & Park, K. (2015). The Influence of Biomedical Information and Childhood History on Sentencing. *Behavioral Sciences & the Law*, *33*(6), 815–826. https://doi.org/10.1002/bsl.2199

Kim-Cohen, J., Caspi, A., Taylor, A., Williams, B., Newcombe, R., Craig, I. W., & Moffitt, T. E. (2006). MAOA, Maltreatment, and Gene–Environment Interaction Predicting Children's Mental Health: New Evidence and a Meta-Analysis. *Molecular Psychiatry*, *11*(10), 903–913. https://doi.org/10.1038/sj.mp.4001851

Klein, W. M., & Kunda, Z. (1992). Motivated Person Perception: Constructing Justifications for Desired Beliefs. *Journal of Experimental Social Psychology*, *28*(2), 145–168. https://doi.org/10.1016/0022-1031(92)90036-J

Knight, J. (2009). Are Empiricists Asking the Right Questions About Judicial Decisionmaking? *Duke Law Journal*, *58*(7), 1531–1556.

Koppel, S., Fondacaro, M., & Na, C. (2018). Cast into Doubt: Free Will and the Justification for Punishment. *Behavioral Sciences & the Law*, *36*(4), 490–506.

Kornhauser, L. A. (1984). The Great Image of Authority. *Stanford Law Review*,

*36*(1/2), 349–389. https://doi.org/10.2307/1228686

Kornhauser, L. A. (1992). Modeling Collegial Courts I: Path-Dependence. *International Review of Law and Economics*, *12*(2), 169–185. https://doi.org/10.1016/0144-8188(92)90034-O

Kornhauser, L. A. (2014). The Economic Analysis of Law. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014). http://plato.stanford.edu/archives/spr2014/entries/legal-econanalysis/

Korobkin, R. B. (2000). A Multi-Disciplinary Approach to Legal Scholarship: Economics, Behavioral Economics, and Evolutionary Psychology. *Jurimetrics*, *41*, 319–336.

Korobkin, R. B. (2004). Possibility and Plausibility in Law and Economics. *Florida State University Law Review*, *32*, 781–795.

Korobkin, R. B. (2006). The Problems with Heuristics for Law. In G. Gigerenzer (Ed.), *Heuristics and the Law* (p. 480). MIT Press.

Korobkin, R. B., & Ulen, T. S. (2000). Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics. *California Law Review*, *88*(4), 1051. https://doi.org/10.2307/3481255

Kuhn, D., Weinstock, M., & Flaton, R. (1994). How Well Do Jurors Reason? Competence Dimensions of Individual Variation in a Juror Reasoning Task. *Psychological Science (0956-7976)*, *5*(5), 289–296. https://doi.org/10.1111/j.1467-9280.1994.tb00628.x

Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Kysar, D., A. (2006). Group Report: Are Heuristics a Problem or a Solution? In C. Engel & G. Gigerenzer (Eds.), *Heuristics And the Law*. MIT Press.

LaDuke, C., Locklair, B., & Heilbrun, K. (2018). Neuroscientific, Neuropsychological, and Psychological Evidence Comparably Impact Legal Decision Making: Implications for Experts and Legal Practitioners. *Journal of Forensic Psychology Research and Practice*, *18*(2), 114–142. https://doi.org/10.1080/24732850.2018.1439142

Lagnado, D. A. (2021). *Explaining the Evidence: How the Mind Investigates the World*. Cambridge University Press.

Lagnado, D. A., & Gerstenberg, T. (2017). Causation in Legal and Moral Reasoning.

In M. R. Waldmann (Ed.), *Oxford Handbook of Causal Reasoning*. Oxford University Press.

Landsman, S., & Rakos, R. F. (1994). A Preliminary Inquiry into the Effect of Potentially Biasing Information on Judges and Jurors in Civil Litigation. *Behavioral Sciences & the Law*, *12*(2), 113–126. https://doi.org/10.1002/bsl.2370120203

Lanteri, A., Chelini, C., & Rizzello, S. (2008). An Experimental Investigation of Emotions and Reasoning in the Trolley Problem. *Journal of Business Ethics*, *83*(4), 789–804. https://doi.org/10.1007/s10551-008-9665-8

Leibovitch, A. (2016). Relative Judgments. *The Journal of Legal Studies*, *45*(2), 281–330. https://doi.org/10.1086/687376

Leiter, B. (1996). Heidegger and the Theory of Adjudication. *The Yale Law Journal*, *106*(2), 253–282. https://doi.org/10.2307/797211

Leiter, B. (1997). Rethinking Legal Realism: Toward a naturalized jurisprudence. *Texas Law Review*, *76*(2), 267.

Leiter, B. (1999). Legal Realism. In D. Patterson (Ed.), *The Blackwell Guide to the Philosophy of Law and Legal Theory* (pp. 261–279). Blackwell Publishers.

Leiter, B. (2001). Legal Realism and Legal Positivism Reconsidered. *Ethics*, *111*(2), 278–301. https://doi.org/10.1086/et.2001.111.issue-2

Leiter, B. (2003). Beyond the Hart/Dworkin Debate: The Methodology Problem in Jurisprudence. *American Journal of Jurisprudence*, *48*, 17.

Leiter, B. (2012). Naturalism in Legal Philosophy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2012). http://plato.stanford.edu/archives/fall2012/entries/lawphil-naturalism/

Levett, L. M., & Devine, D. (2017). Integrating Individual and Group Models of Juror Decision Making. In *The Psychology of Juries* (pp. 11–36). American Psychological Association. https://doi.org/10.1037/0000026-002

Liao, S. M., Wiegmann, A., Alexander, J., & Vong, G. (2012). Putting the Trolley in Order: Experimental Philosophy and the Loop Case. *Philosophical Psychology*, *25*(5), 661–671. https://doi.org/10.1080/09515089.2011.627536

Lieder, F., & Griffiths, T. L. (2019). Resource-Rational Analysis: Understanding Human Cognition as the Optimal Use of Limited Computational Resources. *Behavioral and*

*Brain Sciences*, 1–85. https://doi.org/10.1017/S0140525X1900061X

Lim, Y. (2000). An Empirical Analysis of Supreme Court Justices' Decision Making. *The Journal of Legal Studies*, *29*(2), 721–752. https://doi.org/10.1086/468091

Lindquist, S. A., & Cross, F. B. (2005). Empirically Testing Dworkin's Chain Novel Theory: Studying the Path of Precedent. *New York University Law Review*, *80*(4), 1156–1206.

Liu, J. Z. (2018). Does Reason Writing Reduce Decision Bias? Experimental Evidence from Judges in China. *The Journal of Legal Studies*, *47*(1), 83–118. https://doi.org/10.1086/696879

Liu, J. Z., & Li, X. (2019). Legal Techniques for Rationalizing Biased Judicial Decisions: Evidence from Experiments with Real Judges. *Journal of Empirical Legal Studies*, *16*(3), 630–670. https://doi.org/10.1111/jels.12229

Llewellyn, K. N. (1930). A Realistic Jurisprudence—The Next Step. *Columbia Law Review*, *30*(4), 431–465. https://doi.org/10.2307/1114548

Llewellyn, K. N. (1931). Some Realism about Realism: Responding to Dean Pound. *Harvard Law Review*, *44*(8), 1222–1264. https://doi.org/10.2307/1332182

Lochner v New York, 198 US 45 ___ (1905).

Lombrozo, T. (2009). The Role of Moral Commitments in Moral Judgment. *Cognitive Science*, *33*(2), 273–286. https://doi.org/10.1111/j.1551-6709.2009.01013.x

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.

Lui, J. H. L., Reiter, S. R., Barry, C. T., & Robinson, S. (2019). Effects of Genetic and Environmental Explanations of Psychopathy and Gender on Perceptions of Criminal Behaviors. *The Journal of Forensic Psychiatry & Psychology*, *30*(3), 467–483. https://doi.org/10.1080/14789949.2019.1570542

Lynch, J. M., Lane, J. D., Berryessa, C. M., & Rottman, J. (2019). How Information About Perpetrators' Nature and Nurture Influences Assessments of Their Character, Mental States, and Deserved Punishment. *PLoS One*, *14*(10), e0224093. http://dx.doi.org/10.1371/journal.pone.0224093

Malouff, J., & Schutte, N. S. (1989). Shaping Juror Attitudes: Effects of Requesting Different Damage Amounts in Personal Injury Trials. *Journal of Social Psychology*, *129*(4), 491. https://doi.org/10.1080/00224545.1989.9712067

Marr, D. (2010). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. MIT Press.

Marshall, J., Lilienfeld, S. O., Mayberg, H., & Clark, S. E. (2017). The Role of Neurological and Psychological Explanations in Legal Judgments of Psychopathic Wrongdoers. *The Journal of Forensic Psychiatry & Psychology*, *28*(3), 412–436. https://doi.org/10.1080/14789949.2017.1291706

Marti, M. W., & Wissler, R. L. (2000). Be Careful What You Ask for: The Effect of Anchors on Personal-Injury Damages Awards. *Journal of Experimental Psychology: Applied*, *6*(2), 91–103. https://doi.org/10.1037/1076-898X.6.2.91

Martin, M. (1997). *Legal Realism: American and Scandinavian*. P. Lang.

Maveety, N. L. (Ed.). (2003). *The Pioneers of Judicial Behavior*. University of Michigan Press. https://muse.jhu.edu/book/6364

McCoy, M. L., Nunez, N., & Dammeyer, M. M. (1999). The Effect of Jury Deliberations on Jurors' Reasoning Skills. *Law and Human Behavior*, *23*(5), 557–575.

Mcswiggan, S., Elger, B., & Appelbaum, P. S. (2017). The Forensic Use of Behavioral Genetics in Criminal Proceedings: Case of the MAOA-L Genotype. *International Journal of Law and Psychiatry*, *50*, 17–23. https://doi.org/10.1016/j.ijlp.2016.09.005

Mercier, H. (2010). The Social Origins of Folk Epistemology. *Review of Philosophy and Psychology*, *1*(4), 499–514. https://doi.org/10.1007/s13164-010-0021-4

Mercier, H., & Sperber, D. (2009). Intuitive and Reflective Inferences. In K. Frankish & J. St. B. T. Evans (Eds.), *In Two Minds: Dual Processes and Beyond*. Oxford University Press.

Mercier, H., & Sperber, D. (2011). Why Do Humans Reason? Arguments for an Argumentative Theory. *Behavioral and Brain Sciences*, *34*(02), 57–74.

Mikhail, J. M. (2009). Moral Grammar and Intuitive Jurisprudence: A Formal Model of Unconscious Moral and Legal Knowledge. In Brian H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 50, pp. 27–100). Academic Press.

Mikhail, J. M. (2011). *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge University Press.

Miller, J. R. (2017, July 24). Man Told 911 Before Suicide to 'Hurry up' Because He's an Organ Donor. *New York Post*. https://nypost.com/2017/07/24/man-told-911-before-suicide-to-hurry-up-because-hes-an-organ-donor/

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who Shalt Not Kill? Individual Differences in Working Memory Capacity, Executive Control, and Moral Judgment. *Psychological Science*, *19*(6), 549–557. https://doi.org/10.1111/j.1467-9280.2008.02122.x

Moore, G. E. (1912). *Ethics*. Clarendon Press.

Moore, M. S. (1980). The Semantics of Judging. *Southern California Law Review*, *54*, 151–294.

Moore, U., & Hope, T. S., Jr. (1929). An Institutional Approach to the Law of Commercial Banking. *The Yale Law Journal*, *38*(6), 703–719. https://doi.org/10.2307/790071

Morse, S. J. (2004). New Neuroscience, Old Problems. In B. Garland (Ed.), *Neuroscience and the Law: Brain,Mind,and the Scales of Justice* (2nd edition). Chicago University Press.

Mowle, E. N., Edens, J. F., Clark, J. W., & Sorman, K. (2016). Effects of Mental Health and Neuroscience Evidence on Juror Perceptions of a Criminal Defendant: The Moderating Role of Political Orientation. *Behavioral Sciences & the Law*, *34*(6), 726–741.

Muir, B. R. (2019). *The Influence of Genetic Information and Crime-Type on Juror Decision Making* [Honours, University of Tasmania]. https://eprints.utas.edu.au/34771/

Mussweiler, T., & Strack, F. (1999). Comparing Is Believing: A Selective Accessibility Model of Judgmental Anchoring. *European Review of Social Psychology*, *10*(1), 135–167. https://doi.org/10.1080/14792779943000044

Myers, M. A. (1978). Rule Departures and Making Law: Juries and Their Verdicts. *Law & Society Review*, *13*(3), 781–798.

Nadler, J. (2012). Blaming as a Social Process: The Influence of Character and Moral Emotion on Blame Adjudicating the Guilty Mind. *Law and Contemporary Problems*, *75*(2), 1–32.

Newell, B. R., & Shanks, D. R. (2014). Unconscious Influences on Decision Making:

A Critical Review. *Behavioral and Brain Sciences*, *37*(01), 1–19. https://doi.org/10.1017/S0140525X12003214

Nichols, S., & Mallon, R. (2006). Moral Dilemmas and Moral Rules. *Cognition*, *100*(3), 530–542. https://doi.org/10.1016/j.cognition.2005.07.005

Norcross, A. (2008). Off Her Trolley? Frances Kamm and the Metaphysics of Morality. *Utilitas*, *20*(1), 65–80. https://doi.org/10.1017/S0953820807002919

Northcraft, G. B., & Neale, M. A. (1987). Experts, Amateurs, and Real Estate: An Anchoring-and-Adjustment Perspective on Property Pricing Decisions. *Organizational Behavior and Human Decision Processes*, *39*(1), 84–97. https://doi.org/10.1016/0749-5978(87)90046-X

Nucci, E. D. (2013). Self-Sacrifice and the Trolley Problem. *Philosophical Psychology*, *26*(5), 662–672. https://doi.org/10.1080/09515089.2012.674664

Oaksford, M., & Chater, N. (1998). An Introduction to Rational Models of Cognition. In M. Oaksford & N. Chater (Eds.), *Rational Models of Cognition* (pp. 1–18). Oxford University Press.

Okasha, S. (2002). *Philosophy of Science: A Very Short Introduction*. OUP Oxford.

Oliphant, H. (1926). Return to Stare Decisis, A. *American Law School Review*, *6*, 215.

O'Neill, T. J. (1981). The Language of Equality in a Constitutional Order. *The American Political Science Review*, *75*(3), 626–635. https://doi.org/10.2307/1960957

Ortony, A. (Ed.). (1993). The Role of Similarity in Similes and Metaphors. In *Metaphor and Thought* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139173865

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd edition). Cambridge University Press.

Pennington, D. C. (1982). Witnesses and Their Testimony: Effects of Ordering on Juror Verdicts1. *Journal of Applied Social Psychology*, *12*(4), 318–333. https://doi.org/10.1111/j.1559-1816.1982.tb00868.x

Pennington, N., & Hastie, R. (1988). Explanation-Based Decision Making: Effects of Memory Structure on Judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 521–533. https://doi.org/10.1037/0278-7393.14.3.521

Pennington, N., & Hastie, R. (1991). Cognitive Theory of Juror Decision Making: The Story Model, A. *Cardozo Law Review*, *13*, 519.

Pennington, N., & Hastie, R. (1992). Explaining the Evidence: Tests of the Story Model for Juror Decision Making. *Journal of Personality and Social Psychology*, *62*(2), 189–206. https://doi.org/10.1037/0022-3514.62.2.189

Pennington, N., & Reid, H. (1993). The Story Model for Juror Decision Making. In R. Hastie (Ed.), *Inside the Juror: The Psychology of Juror Decision Making*. Cambridge University Press. http://ebooks.cambridge.org/ref/id/CBO9780511752896

Peresie, J. L. (2004). Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts. *Yale Law Journal*, *114*(7), 1759–1790.

Petrinovich, L., & O'Neill, P. (1996). Influence of Wording and Framing Effects on Moral Intuitions. *Ethology and Sociobiology*, *17*(3), 145–171. https://doi.org/10.1016/0162-3095(96)00041-6

Pinello, D. R. (1998). Linking Party to Judicial Ideology in American Courts: A Meta-Analysis. *Justice System Journal*, *20*(3), 219–254.

Pinker, S. (2009). *How the Mind Works*. W.W. Norton.

Posner, R. A. (1983). *The Economics of Justice* (Reprint edition). Harvard University Press.

Posner, R. A. (1987). The Decline of Law as an Autonomous Discipline: 1962-1987. *Harvard Law Review*, *100*(4), 761–780. https://doi.org/10.2307/1341093

Posner, R. A. (1995). *Overcoming Law*. Harvard University Press.

Posner, R. A. (1998). Rational Choice, Behavioral Economics, and the Law. *Stanford Law Review*, *50*(5), 1551–1575. https://doi.org/10.2307/1229305

Posner, R. A. (2007a). *Economic Analysis of Law* (7th ed..). Aspen Publishers.

Posner, R. A. (2007b). In Memoriam: Bernard D. Meltzer (1914-2007). *University of Chicago Law Review*, *74*, 435.

Posner, R. A. (2008). *How Judges Think*. Harvard University Press.

Pound, R. (1910). Law in Books and Law in Action. *American Law Review*, *44*, 12.

Prakken, H., & Sartor, G. (2012). *Logical Models of Legal Argumentation*. Springer Science & Business Media.

Pritchett, C. H. (1941). Divisions of Opinion Among Justices of the U. S. Supreme Court, 1939-1941. *The American Political Science Review*, *35*(5), 890–898. JSTOR. https://doi.org/10.2307/1948251

Pritchett, C. H. (1948). *The Roosevelt Court: A Study in Judicial Politics and Values, 1937-1947*. Macmillan Co.

Pylyshyn, Z. W. (Ed.). (1987). Preface. In *The Robot's Dilemma: Frame Problem in Artificial Intelligence* (pp. vii–xi). Ablex Publishing Corporation.

Pylyshyn, Z. W. (1999). What's in Your Mind? In E. LePore & Z. W. Pylyshyn (Eds.), *What Is Cognitive Science?* (pp. 1–25). Blackwell.

Quine, W. V. O. (1951). Main Trends in Recent Philosophy: Two Dogmas of Empiricism. *The Philosophical Review*, *60*(1), 20–43.

Quine, W. V. O. (1960). *Word and Object*. Massachusetts Institute of Technology.

Rachlinski, J. J., & Wistrich, A. J. (2017). Judging the Judiciary by the Numbers: Empirical Research on Judges. *Annual Review of Law and Social Science*, *13*(1), 203–229. https://doi.org/10.1146/annurev-lawsocsci-110615-085032

Rachlinski, J. J., Wistrich, A. J., & Guthrie, C. (2017). Judicial Politics and Decisionmaking: A New Approach. *Vanderbilt Law Review*, *70*(6), [i]-2104.

Redding, R. E., & Reppucci, N. D. (1999). Effects of Lawyers' Socio-Political Attitudes on Their Judgments of Social Science in Legal Decision Making. *Law and Human Behavior*, *23*(1), 31–54. JSTOR.

Remmel, R. J., Glenn, A. L., & Cox, J. (2019). Biological Evidence Regarding Psychopathy Does Not Affect Mock Jury Sentencing. *Journal of Personality Disorders*, *33*(2), 164–184. https://doi.org/10.1521/pedi_2018_32_337

Rini, R. A. (2015). How Not to Test for Philosophical Expertise. *Synthese; Dordrecht*, *192*(2), 431–452. http://dx.doi.org.libproxy.ucl.ac.uk/10.1007/s11229-014-0579-y

Robbennolt, J. K., MacCoun, R. J., & Darley, J. M. (2010). Multiple Parallel Constraint Satisfaction. In D. E. Klein & J. D. Gregory Mitchell (Eds.), *The Psychology of Judicial Decision Making*. Oxford University Press.

Robbins, P., & Litton, P. (2018). Crime, Punishment, and Causation: The Effect of Etiological Information on the Perception of Moral Agency. *Psychology, Public Policy, and*

*Law*, *24*(1), 118–127. https://doi.org/10.1037/law0000146

Ross, A. (1946). *Towards a Realistic Jurisprudence; a Criticism of the Dualism in Law*. EMunksgaard.

Rowland, C. K., & Carp, R. A. (1996). *Politics and Judgment in Federal District Courts*. University Press of Kansas.

Rubin, P. H. (2000). Judge-Made Law. *Encyclopedia of Law and Economics*, *5*, 543–558.

Russell, S. J., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (3rd ed.). Pearson.

Saks, M. J., & Kidd, R. F. (1980). Human Information Processing and Adjudication: Trial by Heuristics. *Law & Society Review*, *15*(1), 123–160. https://doi.org/10.2307/3053225

Saks, M. J., Schwieitzer, N. J., Aharoni, E., & Kiehl, K. A. (2014). The Impact of Neuroimages in the Sentencing Phase of Capital Trials. *Journal of Empirical Legal Studies*, *11*(1), 105–300.

Sandys, M., Pruss, H. C., & Walsh, S. M. (2009). Aggravation and Mitigation: Findings and Implications. *Journal of Psychiatry and Law*, *37*(2), 189–236.

Schauer, F. (1988). Formalism. *The Yale Law Journal*, *97*(4), 509–548. https://doi.org/10.2307/796369

Schauer, F. (1995). Giving Reasons. *Stanford Law Review*, *47*(4), 633–659. https://doi.org/10.2307/1229080

Schauer, F. (2010). Is there a Psychology of Judging? In D. E. Klein & J. D. Gregory Mitchell (Eds.), *The Psychology of Judicial Decision Making*. Oxford University Press.

Schauer, F., & Spellman, B. A. (2017). Analogy, Expertise, and Experience. *University of Chicago Law Review*, *84*, 249–268.

Schubert, G. A. (1962). The 1960 Term of the Supreme Court: A Psychological Analysis. *The American Political Science Review*, *56*(1), 90–107. JSTOR. https://doi.org/10.2307/1953099

Schubert, G. A. (1965). *The Judicial Mind: The Attitudes and Ideologies of Supreme Court Justices, 1946-1963*. Northwestern University Press.

Schum, D. A., & Martin, A. W. (1982). Formal and Empirical Research on Cascaded

Inference in Jurisprudence. *Law & Society Review*, *17*(1), 105–151. https://doi.org/10.2307/3053534

Schweitzer, N. J., & Saks, M. J. (2011). Neuroimage Evidence and the Insanity Defense. *Behavioral Sciences & the Law*, *29*(4), 592–607.

Schwitzgebel, E., & Cushman, F. (2012). Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers. *Mind & Language*, *27*(2), 135–153. https://doi.org/10.1111/j.1468-0017.2012.01438.x

Schwitzgebel, E., & Cushman, F. (2015). Philosophers' Biased Judgments Persist Despite Training, Expertise and Reflection. *Cognition*, *141*(Supplement C), 127–137. https://doi.org/10.1016/j.cognition.2015.04.015

Scurich, N., & Appelbaum, P. (2015). The Blunt-Edged Sword: Genetic Explanations of Misbehavior Neither Mitigate nor Aggravate Punishment. *Journal of Law and the Biosciences*, *3*(1), 140–157. https://doi.org/10.1093/jlb/lsv053

Scurich, N., & Appelbaum, P. S. (2017). Behavioural Genetics in Criminal Court. *Nature Human Behaviour*, *1*(11), 772–774. https://doi.org/10.1038/s41562-017-0212-4

Segal, J. A. (1984). Predicting Supreme Court Cases Probabilistically: The Search and Seizure Cases, 1962-1981. *The American Political Science Review*, *78*(4), 891–900. https://doi.org/10.2307/1955796

Segal, J. A., & Cover, A. D. (1989). Ideological Values and the Votes of U.S. Supreme Court Justices. *The American Political Science Review*, *83*(2), 557–565. JSTOR. https://doi.org/10.2307/1962405

Segal, J. A., & Spaeth, H. J. (1996a). Norms, Dragons, and Stare Decisis: A Response. *American Journal of Political Science*, *40*(4), 1064–1082. JSTOR. https://doi.org/10.2307/2111743

Segal, J. A., & Spaeth, H. J. (1996b). The Influence of Stare Decisis on the Votes of United States Supreme Court Justices. *American Journal of Political Science*, *40*(4), 971–1003. JSTOR. https://doi.org/10.2307/2111738

Segal, J. A., & Spaeth, H. J. (2002). *The Supreme Court and the Attitudinal Model Revisited*. Cambridge University Press. https://doi.org/10.1017/CBO9780511615696

Shapiro, D. L. (1986). In Defense of Judicial Candor. *Harvard Law Review*, *100*(4),

731–750.

Shapiro, M. (1972). Toward a Theory of Stare Decisis. *Journal of Legal Studies*, *1*(1), 125–134.

Shaw, D. M. (2014). Organ Donation After Assisted Suicide: A Potential Solution to the Organ Scarcity Problem. *Transplantation*, *98*(3), 247–251. https://doi.org/10.1097/TP.0000000000000099

Sheehan, R. S., Mishler, W., & Songer, D. R. (1992). Ideology, Status, and the Differential Success of Direct Parties Before the Supreme Court. *The American Political Science Review*, *86*(2), 464–471. JSTOR. https://doi.org/10.2307/1964234

Simon, D. (1998). Psychological Model of Judicial Decision Making, A. *Rutgers Law Journal*, *30*, 1.

Simon, D. (2004). A Third View of the Black Box: Cognitive Coherence in Legal Decision Making. *The University of Chicago Law Review*, *71*(2), 511–586.

Simon, D. (2010). In Praise of Pedantic Eclecticism: Pitfalls and Opportunities in the Psychology of Judging. In D. E. Klein & G. Mitchell (Eds.), *The Psychology of Judicial Decision Making*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367584.001.0001

Simon, D., Krawczyk, D. C., & Holyoak, K. J. (2004). Construction of Preferences by Constraint Satisfaction. *Psychological Science (0956-7976)*, *15*(5), 331–336.

Simon, D., Pham, L. B., Le, Q. A., & Holyoak, K. J. (2001). The Emergence of Coherence Over the Course of Decision Making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1250.

Simon, D., Snow, C. J., & Read, S. J. (2004). The Redux of Cognitive Consistency Theories: Evidence Judgments by Constraint Satisfaction. *Journal of Personality and Social Psychology*, *86*(6), 814–837. https://doi.org/10.1037/0022-3514.86.6.814

Simon, D., & Spiller, S. A. (2016). The Elasticity of Preferences. *Psychological Science*, *27*(12), 1588–1599. https://doi.org/10.1177/0956797616666501

Simon, H. A. (1956). Rational Choice and the Structure of the Environment. *Psychological Review*, *63*(2), 129–138. https://doi.org/10.1037/h0042769

Simon, H. A. (1992). What Is an "Explanation" of Behavior? *Psychological Science*,

*3*(3), 150–161. https://doi.org/10.1111/j.1467-9280.1992.tb00017.x

Sinnott-Armstrong, W. (2007). Framing Moral Intutions. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 2, The Cognitive Science of Morality: Intutition and Diversity*. The MIT Press. http://cognet.mit.edu/book/moral-psychology-0

Sisk, G. C., & Heise, M. (2004). Judges and Ideology: Public and Academic Debates about Statistical Measures. *Northwestern University Law Review*, *99*(2), 743–804.

Sloman, S. A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin January 1996*, *119*(1), 3–22.

Smart, J. J. C. (1961). Free-Will, Praise and Blame. *Mind*, *70*(279), 291–306. JSTOR.

Smolensky, P. (1988). On the Proper Treatment of Connectionism. *Behavioral and Brain Sciences*, *11*(1), 1–23. https://doi.org/10.1017/S0140525X00052432

Snyder, M. L., Kleck, R. E., Strenta, A., & Mentzer, S. J. (1979). Avoidance of the Handicapped: An Attributional Ambiguity Analysis. *Journal of Personality and Social Psychology*, *37*(12), 2297–2306.

Sood, A. M. (2013). Motivated Cognition in Legal Judgments—An Analytic Review. *Annual Review of Law and Social Science*, *9*(1), 307–325. https://doi.org/10.1146/annurev-lawsocsci-102612-134023

Sood, A. M., & Darley, J. M. (2012). The Plasticity of Harm in the Service of Criminalization Goals. *California Law Review*, *100*(5), 1313–1358.

Spaeth, H. J. (1961). An Approach to the Study of Attitudinal Differences as an Aspect of Judicial Behavior. *Midwest Journal of Political Science*, *5*(2), 165–180. JSTOR. https://doi.org/10.2307/2109268

Spamann, H., & Klöhn, L. (2016). Justice Is Less Blind, and Less Legalistic, than We Thought: Evidence from an Experiment with Real Judges. *The Journal of Legal Studies*. https://doi.org/10.1086/688861

Spamann, H., Klöhn, L., Jamin, C., Khanna, V., Liu, J. Z., Mamidi, P., Morell, A., & Reidel, I. (2021). Judges in the Lab: No Precedent Effects, No Common/Civil Law Differences. *Journal of Legal Analysis*, *13*(1), 110–126. https://doi.org/10.1093/jla/laaa008

Spellman, B. A. (2010). Judges, Expertise, and Analogy. In D. E. Klein & J. D. Gregory Mitchell (Eds.), *The Psychology of Judicial Decision Making* (pp. 149–164). Oxford

University Press.

Spellman, B. A., Ullman, J. B., & Holyoak, K. J. (1993). A Coherence Model of Cognitive Consistency: Dynamics of Attitude Change During the Persian Gulf War. *Journal of Social Issues*, *49*(4), 147–165. https://doi.org/10.1111/j.1540-4560.1993.tb01185.x

Sperber, D. (2001). An Evolutionary Perspective on Testimony and Argumentation. *Philosophical Topics*, *29*(1 & 2), 177–189.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, *25*(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Sperber, D., & Mercier, H. (2012). Reasoning as a Social Competence. In H. Landemore & J. Elster (Eds.), *Collective Wisdom: Principles and Mechanisms*. Cambridge University Press.

Spottswood, M. (2013). Bridging the Gap Between Bayesian and Story-Comparison Models of Juridical Inference. *Law, Probability and Risk*, mgt010. https://doi.org/10.1093/lpr/mgt010

Stintzing, R. (1857). *Ulrich Zasius: Ein Beitrag zur Geschichte der Rechtswissenschaft im Zeitalter der Reformation*. Schweighauser.

Strawson, P. F. (1992). *Analysis and Metaphysics: An Introduction to Philosophy*. OUP Oxford.

Suicide Act, Pub. L. No. c. 60 (1962).

Sunstein, C. R. (1995). Incompletely Theorized Agreements: Harvard Law Review. *Harvard Law Review*, *108*(7), 1733. https://doi.org/10.2307/1341816

Sunstein, C. R. (2005). Moral Heuristics. *Behavioral and Brain Sciences*, *28*(04), 531–542. https://doi.org/10.1017/S0140525X05000099

Sunstein, C. R., Schkade, D., Ellman, L. M., & Sawicki, A. (2006). *Are Judges Political?: An Empirical Analysis of the Federal Judiciary*. Brookings Institution Press. https://www.jstor.org/stable/10.7864/j.ctt12879t7

Sunstein, C. R., & Ullmann-Margalit, E. (1999). Second-Order Decisions. *Ethics*, *110*(1), 5–31. https://doi.org/10.1086/233202

Taroni, F., Biedermann, A., Bozza, S., Garbolino, P., & Aitken, C. (2014). *Bayesian*

*Networks for Probabilistic Inference and Decision Analysis in Forensic Science*. John Wiley & Sons, Incorporated.

Tate, C. N. (1981). Personal Attribute Models of the Voting Behavior of U.S. Supreme Court Justices: Liberalism in Civil Liberties and Economics Decisions, 1946-1978. *The American Political Science Review*, *75*(2), 355–367. JSTOR. https://doi.org/10.2307/1961370

Tetlock, P. E. (1983). Accountability and the Perseverance of First Impressions. *Social Psychology Quarterly*, *46*(4), 285–292. https://doi.org/10.2307/3033716

Thagard, P. (1989). Explanatory Coherence. *Behavioral and Brain Sciences*, *12*(03), 435–467. https://doi.org/10.1017/S0140525X00057046

Thagard, P. (2004). Causal Inference in Legal Decision Making: Explanatory Coherence vs. Bayesian Networks. *Applied Artificial Intelligence*, *18*(3–4), 231–249. https://doi.org/10.1080/08839510490279861

Thompson, R. S. (1985). Legitimate and Illegitimate Decisional Inconsistency: A Comment on Brilmayer's Wobble, or the Death of Error. *Southern California Law Review*, *59*, 423.

Thomson, J. J. (1976). Killing, Letting Die, and the Trolley Problem. *The Monist*, *59*(2), 204.

Thomson, J. J. (1995). The Trolley Problem. *Yale Law Journal*, *94*, 1395.

Todd, P. M., & Gigerenzer, G. (1999). Fast and Frugal Heuristics: The Adaptive Toolbox. In G. Gigerenzer & P. M. Todd (Eds.), *Simple Heuristics That Make Us Smart* (pp. 3–34). Oxford University Press.

Toulmin, S. E. (1958). *The Uses of Argument* (2nd ed.). Cambridge University Press.

Truong, T. N., Kelley, S. E., & Edens, J. F. (2021). Does Psychopathy Influence Juror Decision-Making in Capital Murder Trials? "The Devil is in the (Methodological) Details". *Criminal Justice and Behavior*, *48*(5), 690–707. https://doi.org/10.1177/0093854820966369

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

Ulen, T. S. (2000). Rational Choice Theory in Law and Economics. In B. Bouckaert & G. D. Geest (Eds.), *Encyclopedia of Law and Economics: The History and Methodology of*

*Law and Economics v. 1* (pp. 790–818). Edward Elgar Publishing Ltd.

Ulmer, S. S. (1960). Supreme Court Behavior and Civil Rights. *The Western Political Quarterly*, *13*(2), 288–311. JSTOR. https://doi.org/10.2307/444651

Unger, P. K. (1996). *Living High and Letting Die: Our Illusion of Innocence*. Oxford University Press. http://dx.doi.org/10.1093/0195108590.001.0001

Van Raemdonck, D. (2011). Initial Experience with Transplantation of Lungs Recovered from Donors After Euthanasia. *Applied Cardiopulmonary Pathophysiology*, *15*, 38–48.

Van Raemdonck, D., Neyrinck, A., Dupont, L., Coosemans, W., Decaluwé, H., De Leyn, P., Nafteux, P., & Verleden, G. M. (2011). 24 Transplantation of Lungs Recovered from Donors after Euthanasia. *The Journal of Heart and Lung Transplantation*, *30*(4, Supplement), S16. https://doi.org/10.1016/j.healun.2011.01.031

Veljanovski, C. (2006). *The Economics of Law* (2nd ed.). The Institute of Economic Affairs. http://papers.ssrn.com/abstract=935952

Walton, D. N. (2002). *Legal Argumentation and Evidence*. Pennsylvania State University Press.

Walton, D. N. (2005). *Fundamentals of Critical Argumentation*. Cambridge University Press.

Wambaugh, E. (1894). *The Study of Cases: A Course of Instruction in Reading and Stating Reported Cases, Composing Head-Notes and Briefs, Criticising and Comparing Authorities, and Compiling Digests*. Boston: Little, Brown, and Co.

War Office. (1914). *Manual of Military Law* (6th ed.). His Majesty's Stationary Office. https://heinonline.org/HOL/P?h=hein.cow/mailaw0001&i=1

Wicklund, R. A., & Brehm, J. W. (1976). *Perspectives on Cognitive Dissonance* (pp. xiv, 349). Lawrence Erlbaum.

Wiegmann, A., Okan, Y., & Nagel, J. (2012). Order Effects in Moral Judgment. *Philosophical Psychology*, *25*(6), 813–836. https://doi.org/10.1080/09515089.2011.631995

Wiegmann, A., & Waldmann, M. R. (2014). Transfer Effects Between Moral Dilemmas: A Causal Model Theory. *Cognition*, *131*(1), 28–43. https://doi.org/10.1016/j.cognition.2013.12.004

Wigmore, J. H. (1913). *The Principles of Judicial Proof as Given by Logic, Psychology, and General Experience, and Illustrated in Judicial Trials*. Boston : Little, Brown, and Company. http://archive.org/details/principlesofjudi00wigm

Wilkinson, D., & Savulescu, J. (2012). Should We Allow Organ Donation Euthanasia? Alternatives for Maximizing the Number and Quality of Organs for Transplantation. *Bioethics*, *26*(1), 32–48. https://doi.org/10.1111/j.1467-8519.2010.01811.x

Wilson, W. (1971). Source Credibility and Order Effects. *Psychological Reports*, *29*(3_suppl), 1303–1312. https://doi.org/10.2466/pr0.1971.29.3f.1303

Wissler, R. L., & Saks, M. J. (1985). On the Inefficacy of Limiting Instructions. *Law and Human Behavior*, *9*(1), 37–48. https://doi.org/10.1007/BF01044288

Wistrich, A. J., Guthrie, C., & Rachlinski, J. J. (2004). Can Judges Ignore Inadmissible Information—The Difficulty of Deliberately Disregarding. *University of Pennsylvania Law Review*, *153*(4), 1251–1346.

Wistrich, A. J., Rachlinski, J. J., & Guthrie, C. (2015). Heart Versus Head: Do Judges Follow the Law or Follow Their Feelings? *Texas Law Review; Austin*, *93*(4), 855–923.

Ysebaert, D., Van Beeumen, G., De Greef, K., Squifflet, J. P., Detry, O., De Roover, A., Delbouille, M.-H., Van Donink, W., Roeyen, G., Chapelle, T., Bosmans, J.-L., Van Raemdonck, D., Faymonville, M. E., Laureys, S., Lamy, M., & Cras, P. (2009). Organ Procurement After Euthanasia: Belgian Experience. *Transplantation Proceedings*, *41*(2), 585–586. https://doi.org/10.1016/j.transproceed.2008.12.025

Zamir, E., Teichman, D., Teichman, D., & Zamir, E. (2014). Judicial Decision-Making. In E. Zamir & D. Teichman (Eds.), *The Oxford Handbook of Behavioral Economics and the Law*. Oxford University Press. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199945474.001.0001/oxfordhb-9780199945474-e-026