

Northumbria Research Link

Citation: Dargue, Paul (2016) The Impartiality of the England and Wales Court of Appeal (Criminal Division): A Quantitative Analysis. Doctoral thesis, Northumbria University.

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/33939/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>



**Northumbria
University**
NEWCASTLE



UniversityLibrary

The Impartiality of the England and Wales Court of Appeal (Criminal Division): A Quantitative Analysis

P Dargue

PhD

2016

The Impartiality of the England and Wales Court of Appeal (Criminal Division): A Quantitative Analysis

Paul Dargue

A thesis submitted in partial fulfilment of
the requirements of the University of
Northumbria at Newcastle for the
degree of
Doctor of Philosophy

October 2016

Abstract

This thesis analyses the development, methodology, and results of a quantitative study of the decision-making of the England and Wales Court of Appeal (Criminal Division). The Court of Appeal plays an important constitutional role, and the impartiality of the judges is central to its legitimacy. Drawing upon research from the Empirical Legal Studies (ELS) research community, this thesis explores the question of the Court of Appeal's impartiality.

As an incomplete measurement of impartiality, a sample of the Court of Appeal's decisions has been analysed. A dataset of all murder and rape appeals against conviction decided between 2006 and 2010 has been created. A range of factual, demographic, and legal variables have been collected from each of these 472 appeals against conviction, utilising quantitative content analysis. It has been determined, utilising binary logistic regression analysis, whether the variables under analysis are predictors of the outcome of appeals against conviction.

Almost all of the variables analysed showed only a limited ability to predict the outcomes of appeals. Moreover, this study finds support for the legal model of judicial decision-making. A variable designed to capture impartial decision-making had the strongest association with the outcome of appeals. However, a small number of factual and demographic variables are shown to be predictors of outcomes. There is insufficient evidence to doubt the impartiality of the Court of Appeal, but the emergence of these patterns in the data warrants further investigation. This conclusion is important to users and observers of the Court, to whom the impartiality, and so legitimacy, of the Court's decision-making is essential.

Contents

Abstract	3
Contents	4
List of Tables and Figures.....	9
Acknowledgments	11
Author's Declaration.....	12
Chapter 1: Introduction	13
1.1 The aims of this thesis	13
1.2 The importance of impartiality	18
1.3 An analysis of murder and rape appeals against conviction	21
1.4 Thesis overview	22
1.5 Conclusion	23
Chapter 2: The Concept of Impartiality	24
Introduction.....	24
2.1 The constitutional role of judges	25
2.2 The rule of law paradigm and the legal model	27
2.3 What is impartiality?.....	32
2.4 Judicial discretion and impartiality	35
2.5 Independence and impartiality	41
2.6.1 The problematics of subjective impartiality	43
2.6.2 The problematic of diversity.....	44
2.6.3 Psychology and impartiality	49
2.7 Objective impartiality.....	51
2.8 Conclusion.....	55

Chapter 3: Mapping the Empirical Legal Studies Field	57
Introduction.....	57
3.1 An explorative positivistic study of judicial decision-making	58
3.2 Objectivity	62
3.3 American Legal Realism.....	63
3.4 New Legal Realism.....	70
3.5 Behavioural models of judicial decision-making	73
3.6 Post-attitudinal models of judicial decision-making.....	80
3.7 The relationship between this thesis and previous judicial behaviour studies	85
3.8 Conclusion.....	87
Chapter 4: The England and Wales Court of Appeal (Criminal Division)...	89
Introduction.....	89
4.1 Purposes and function of the Court of Appeal	90
4.2 Criminal justice system statistics	95
4.3 A short history of the Court's powers	102
4.4 Analysis of the 'unsafety test'	112
4.5 The meaning of unsafety	116
4.6 Procedural irregularities and unfairness	119
4.7 Fresh evidence appeals.....	124
4.8 Lurking doubt appeals	129
4.9 Conclusion.....	131

Chapter 5: Study Design, Data Collection and Methods of Analysis 132

Introduction.....	132
5.1 Quantitative content analysis.....	132
5.2 Measuring impartiality.....	136
5.3 Impartiality and its ‘observable implications’.....	137
5.4 The role of law in the decision-making process.....	145
5.5 The fundamental problem of causal inference.....	148
5.6 Hypothesis testing, the p-value, and statistical significance.....	151
5.7.1 Parameters of sampled appeals.....	155
5.7.2 Parameters of sampled murder appeals against conviction.....	158
5.7.3 Parameters of sampled rape appeals against conviction.....	160
5.8 Binary logistic regression.....	164
5.9 Conclusion.....	169

Chapter 6: Evaluation of Variables and Methods of Study..... 171

Introduction.....	171
6.1.1 The dependant variable.....	172
6.1.2 Factual and demographic variables.....	174
6.1.3 Grounds of appeal raised.....	182
6.1.4 Legal and institutional variables.....	187
6.2 Data collection template.....	190
6.3.1 Evaluation of data collection.....	198
6.3.2 Evaluation of the collection of grounds of appeal.....	202
6.4 Research strengths and limitations.....	204
6.5 Conclusion.....	211

Chapter 7: Findings and Analysis	212
Introduction.....	212
7.1 Overview of sample profile	213
7.2.1 Grounds of appeal raised in sample of murder and rape appeals	216
7.2.2 Appeals raising fresh evidence as a ground of appeal	221
7.2.3 The ‘Insufficient Evidence, Lurking Doubt, Generally Unsafe’ ground of appeal.....	224
7.3 Summary of findings	227
7.4 Purposive selection procedure	228
7.5 Results of initial multiple binary logistic regression analysis	234
7.6 Binary logistic regression Model 1	238
7.7.1 The law governing the case	244
7.7.2 Discussion	248
7.8 Inclusion of the primary legal variable as an independent variable.....	250
7.9 Summary of findings	252
7.10 Additional analysis of data – fresh evidence and CCRC referrals	253
7.11 Conclusion.....	258
Chapter 8: Discussion: The Impartiality of the England and Wales Court of Appeal (Criminal Division)	260
Introduction.....	260
8.1 Overview of study	261
8.2 Grounds of appeal argued in the Court of Appeal in murder and rape appeals – Is the Court of Appeal restrictive?	264
8.3 Analysing the impartiality of the Court of Appeal	271
8.4 How successful is the study as an exploration of impartiality?	279
8.5 The ‘unsafety test’ and theories of judicial decision-making	284
8.6 Outcome of null hypothesis tests of judicial impartiality	287

8.7 Conclusion.....	289
Chapter 9: Thesis Conclusions	290
Introduction.....	290
9.1 Summary of original contributions.....	291
9.2 Decision-making in the England and Wales Court of Appeal (Criminal Division).....	294
9.3 Strengths and limitations of study.....	295
9.4 Potential avenues for future research	297
9.5 Conclusion.....	298
Appendix A: List of Variables.....	300
Appendix B: List of Sampled Appeals.....	302
Appendix C: Purposive Selection Results.....	304
Appendix D: Binary Logistic Regression Tables: Model of Legal Variables.....	313
Bibliography.....	315

List of Tables and Figures

Table 1.1: Comparison of overall Court of Appeal workload and murder and rape workload (2006-2010).....	22
Table 4.1: Number of defendants convicted in a Crown Court following not guilty plea 2007-2015.	97
Table 4.2: Persons convicted of rape of a female 2006 - 2010.	98
Table 4.3: Persons convicted of rape of a male 2006 - 2010.	98
Table 4.4: Total number of persons convicted of rape 2006 – 2010.....	98
Table 4.5: Number of persons convicted of murder 2006 – 2010.....	99
Table 4.6: Applications received 2006 – 2010.....	100
Table 4.7: Comparison of overall Court of Appeal workload and murder and rape workload (2006-2010).....	101
Table 5.1: Age profile of complainants / deceased in sample.	157
Table 5.2: Appeals in sample by gender.	157
Table 7.1: Outcome of sampled appeals per year, 2006-2010.....	214
Table 7.2: Percentage of murder and rape appeals successful per year, 2006 - 2010.....	215
Table 7.3: Overall success rates in Court of Appeal 2006-2010.....	215
Table 7.4(a): Grounds of appeal raised in sampled appeals: Overall.....	218
Table 7.4(b): Grounds of appeal raised in sampled appeals: Rape.	219
Table 7.5: Categories of fresh evidence.....	222
Table 7.6: Purposive selection procedure results. Cut off: $p = > 0.25$	231
Table 7.7: Classification table: Initial analysis.	235
Table 7.8: Variables in equation: Initial analysis.....	237
Table 7.9: Classification table: Model 1.....	238
Table 7.10: Variables in equation: Model 1.	239
Table 7.11: 95% confidence interval for exp(B): Model 1.....	240
Table 7.12: Number of appeals allowed / dismissed, by whether an error occurred.....	245
Table 7.13: Classification table: Model 2.....	247
Table 7.14: Variables in equation: Model 2.	247
Table 7.15: Classification Table: Model 3.....	250

Table 7.16: Variables in equation: Model 3.	251
Table 7.17: 95% confidence intervals for exp(B): Model 3.	251
Table 7.18: Results of X^2 Analysis.	255
Table 7.19: Phi-coefficients.	256
Table 7.20: Classification table: Model 4.	257
Table 7.21: Variables in equation: Model 4.	257
Figure 6.1: Treatment of successful appeals.	173
Figure 7.1: Outcome of appeals against conviction for murder and rape 2006 - 2010.	214
Figure 7.2: Outcome of sampled rape appeals against conviction.	214
Figure 7.3: Outcome of sampled murder appeals against conviction.	215

Acknowledgments

I give my sincerest thanks to my supervisory team of Professor Raymond Arthur, Dr Carole McCartney, and Dr Andrew Robson, for their patience and guidance. I also thank the numerous other academic and administrative colleagues who have in some way contributed to the completion of this thesis. Special thanks are due to Professor Paul van Schaik (Teesside University); my examiners and mock examiners; and the innumerable scholars whose work and ideas have contributed to this study.

I thank my parents (Sylvia and Phil) for their support. Most of all I thank my wife Alexa, for her endless support and assistance throughout this whole process. Your efforts and faith kept this project going. Finally, I thank our baby Ava. You may not know it yet, but your 3am wake up calls contributed greatly to the completion of this thesis.

For dad.

Author's Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by the University Ethics Committee on 20 May 2014.

I declare that the word count of this thesis is 83710

Name:

Paul Dargue

Signature:

Date:

Chapter 1

Introduction

1.1 The aims of this thesis

This study is a quantitative empirical study of the England and Wales Court of Appeal (Criminal Division).¹ This Court forms part of the State's response to crime, in which it is intended that crimes are detected and investigated, and that the true perpetrators of crimes are convicted in accordance with the rule of law and due process.² The Court hears appeals against Crown Court convictions and is required to quash any conviction which it thinks is 'unsafe'.³ In exercising its function of reviewing the safety of convictions, the Court forms an important part of the mechanism of criminal justice in England and Wales. For several decades, academic studies and governmental reviews of the Court have analysed how it operates its powers.⁴ This study addresses a specific question regarding the Court's decision-making: whether it appeared to have determined appeals against conviction for murder and rape in an impartial manner.

The question of the Court's impartiality is a normative one, but, as Epstein and Martin explained, this question cannot be answered directly because impartiality cannot be directly observed or completely measured.⁵ This thesis utilises the methods of quantitative Empirical Legal Studies (ELS) to address the question of the impartiality of the Court of Appeal in a sample of cases decided by it.⁶ Impartiality is measured by utilising variables which are 'observable implications' of impartial decision-making, or a lack of it. The development of this methodology

¹ Hereafter 'the Court' or 'Court of Appeal', unless otherwise stated.

² See A Ashworth and M Redmayne, *The Criminal Process* (4th ed, Oxford University Press 2010) 2.

³ Criminal Appeal Act 1968, as amended by Criminal Appeal Act 1995, s 2.

⁴ See Royal Commission on Criminal Justice, *Report* (cm 2263 1993) (RCCJ); S Roberts, 'The Royal Commission on Criminal Justice and Factual Innocence: Remediating Wrongful Convictions in the Court of Appeal' (2004) (1)(2) JJ 86; SJ Heaton, *A Critical Evaluation of Using Innocence as a Criterion in the Post-Conviction Process* (DPhil Thesis, University of East Anglia 2013); M Naughton, *The Innocent and the Criminal Justice System: A Sociological Analysis of Miscarriages of Justice* (Palgrave Macmillan 2013).

⁵ L Epstein and AD Martin, *An Introduction to Empirical Legal Research* (Oxford University Press 2014) 29-30.

⁶ See TJ Miles and CR Sunstein, 'The New Legal Realism' (2008) 75 U Chic L Rev 831, who describe judicial studies quantitative ELS as one element of a larger empirical project.

and its application to this Court is one of the major themes of this thesis. The Court's decision-making is analysed empirically in two ways. Firstly, the grounds of appeal raised in the murder and rape appeals are analysed. This seeks to replicate previous studies of the Court. Secondly, binary logistic regression analysis is utilised to determine whether a range of variables are predictors of, or correlated with, the decision of the Court to either allow or dismiss an appeal against conviction. It is this second element of the research which is primarily used to assess the impartiality of the Court.

To be successful in addressing the question of impartiality, it must be shown that the question answered – whether statistical associations or patterns exist in the data – and the normative question raised – impartiality – are sufficiently close.⁷ To achieve this, a range of factual, demographic, and legal variables have been collected from each case included in the study. These are the observable implications which provide a proxy for the principle of impartial decision-making. These data have been collected via quantitative content analysis of the Court's judgments, utilising the aid of a data collection template and a variable coding guide. These data are analysed for the presence of statistical associations between the variables and the outcome of appeals against conviction for the offences of murder and rape, decided between 2006 and 2010. The ultimate aim of this thesis is to commence an exploratory step towards a robust, quantitative, statistical analysis of the decision-making of the Court of Appeal. This study could be considered an explorative data analysis of a unique dataset of Court decisions. Explorative data analysis focuses upon the presence of patterns in the data, to allow for inductive theorising.⁸

⁷ See Epstein and Martin (n 5) Chapter 1.

⁸ AT Jebb, S Parrigon, SE Woo, 'Exploratory Data Analysis as a Foundation of Inductive Research' (2017) HR Man Rev 265.

The branch of ELS known as ‘judicial studies’ has proposed a number of theories of judicial decision-making. The traditional view of judicial decision-making is embodied by the ‘legal model’ – the theory that judges impartially apply the law when resolving disputes.⁹ Against this, some theories, such as the behavioural and attitudinal models postulate that personal characteristics or ideology mould how judges think, and so explain judicial decision-making.¹⁰ Other models have considered the broader institutional context in which courts and judges operate.¹¹ This requires a consideration of the ‘rules that structure social interactions’, in particular formal laws and informal norms of judicial behaviour.¹² These models are discussed in Chapter 3 of this thesis, and variables from these models have been drawn upon in this study. Given its explorative nature, this study does not seek to test these models fully, but seeks to make use of the data to explore the presence of patterns.

Most ELS research on judicial decision-making is American, and these models, and others, have been tested for several decades.¹³ Such studies are rare in Britain, and, indeed, most countries outside of the United States.¹⁴ One possible reason for this is that American Legal Realism had a far more limited influence on the British legal academy, when compared to the US legal academy.¹⁵ A seminal 1897 article, by American judge and scholar Oliver Wendell Holmes Jr, could be considered the starting-point of the Legal Realism movement. He wrote:

‘If we take our friend the bad man, we shall find that he does not care two straws for the axioms or deductions ... [he wants] to know what the Massachusetts or the English courts are likely to do in fact. I am much of this mind. The prophesies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law’.¹⁶

⁹ See FB Cross, ‘Decisionmaking in the US Circuits Courts of Appeals’ (2003) 91 Cal L Rev 1459.

¹⁰ NL Maveety (Ed), *The Pioneers of Judicial Behavior* (University of Michigan Press 2003).

¹¹ L Epstein and J Knight, *The Choices Justices Make* (CQ Press 1998) 17.

¹² *ibid.*

¹³ See M Maveety ‘The Study of Judicial Behavior’ in Maveety (ed) (n 10).

¹⁴ See Z Robinson, ‘Comparative Judicial Attitudinalism: A Preliminary Study of Judicial Choices in Westminster Legal Systems’ (2011) U Chi Legal F 209, 209.

¹⁵ See M Adler and J Simon, ‘Stepwise Progression: The Past, Present, and Possible Future of Empirical Research on Law in the United States and United Kingdom’ (2014) J Law & Soc 173.

¹⁶ OW Holmes, ‘Path of the Law’ (1897) 1 Boston L School Mag 1, 4.

In order to discover how courts would in fact decide, Holmes said, 'the black-letter man may be the man of the present, but the man of the future is the man of statistics'.¹⁷ Whilst the use of statistics to test theories of judicial decision-making is well entrenched in the American legal academy, the method is largely in its infancy in Britain. This thesis seeks to begin to redress this misbalance, by completing a quantitative empirical study of judicial decision-making. No similar study of the relationship between a range of variables and decisions in the Court of Appeal exists. By conducting this study, it is intended that a methodological contribution will be made, by demonstrating the strengths and weaknesses of this approach and its application to the Court of Appeal.

A list of variables used in this study is provided in Appendix A. In addition, a full explanation of all the variables which are analysed in this study is provided in Chapter 6. Readers who wish an early review of the variables used in this study should turn to Appendix A. Epstein and King proposed that in conducting quantitative analysis researchers should, 1) invoke theories that produce *observable implications*, 2) extract as many implications as possible, and 3) delineate how they plan to observe those implications.¹⁸ It is explained throughout this thesis how it has been sought to adhere to these principles, and it is evaluated how successful this has been. In order to seek to extract as many implications as possible, data has been collected from each case in the sample in the following areas:

- The bench; (individual judge, genders, ranks, etc.);
- The appellant (gender, age, previous convictions etc.);
- The deceased / complainant (gender, age range etc.);
- The trial (unanimity, the trial judge, etc.);
- Kind of offending (historical, weapons, known / stranger etc.);
- Grounds of appeal (grounds raised, how they were dealt with by the court etc.);

¹⁷ *ibid*, 11.

¹⁸ See L Epstein and G King, 'The Rules of Inference' (2002) 69 U Chi L Rev 1, 47 (emphasis added).

- Sentencing;
- A variable capturing the law

These data were then used to develop the variables which are the observable implications of impartial decision-making, or a lack of impartiality. Appendix A shows the variables categorised into factual and demographic variables, institutional variables, and legal variables.

Whilst the collection of these variables allows for some close analysis of the Court's decision-making, it has limitations. This study takes the form of a non-reactive, non-experimental, observational study. It is a database study, which seeks to determine the presence or absence of statistical relationships between variables and judicial decisions. There are weaknesses to this approach,¹⁹ which are explored fully in this thesis. To be able to capture impartiality completely, all data relating to both knowable and *unknowable* factors would need to be collected and analysed. Such an ideal is impossible even for knowable factors, because it is not possible to control all potential explanatory variables.²⁰ Thus, the variables collected in this study cannot completely capture the principle of impartiality – impartiality cannot be completely validly captured.

A key question addressed in this thesis is which of the particular selected variables are the strongest predictors (i.e. have the strongest relationship with) the outcome of appeals against conviction. Some of these variables, in particular those drawn from the behavioural model and its successor the attitudinal model, may be observable implications of a lack of impartiality, as they are not legally relevant variables. Other variables, indicative of the legal model, may be observable implications of impartial decision-making. This gives an initial exploration of whether the data are consistent with the Court having determined appeals in an impartial manner.

¹⁹ See HT Edwards and MA Livermore, 'Pitfalls of Empirical Studies that Attempt to Understand the Factors Affecting Appellate Decision Making' (2009) (58) Duke LJ 1895

²⁰ See Epstein and Martin (n 5 above) at 7-10.

In order to place this study within its proper epistemological context, the results must be interpreted cautiously. It is only possible to report the presence, or absence, of statistical patterns between the particular variables under analysis and the outcome of the appeals in the sample. There is a gap between the variables actually collected and the variables which ideally would be collected in order to fully assess the impartiality of the Court. In summary, this means that the presence of associations or patterns between variables and outcomes does not prove that the Court lacked impartiality. Conversely, the absence of associations does not mean that the Court did decide appeals in an impartial manner. Rather, the emergence of patterns in the data, either association or lack of association between variables and outcomes, will give material for further exploration and induction of the issues uncovered, and for the development of further hypotheses. The study must therefore be understood as an explorative and inductive study of a range of variables which could potentially be associated with the outcome of appeals, and so a first step towards analysing decision-making of the Court in this manner.

1.2 The importance of impartiality

The normative question addressed in this thesis is whether the Court appeared to have determined appeals against conviction in an impartial manner. As explained above, this is addressed in an incomplete manner by the collection of data from Court of Appeal judgments which are then analysed statistically. Impartiality is required whenever 'there exists a conflict of interests between two or more parties, with a third party being involved to either police the conflict or to resolve it'.²¹ In the Court of Appeal, there is usually a dispute between the Crown and an appellant. The judges in the Court of Appeal are called upon to resolve such disputes, and so they must resolve them in an impartial manner. Impartiality requires that the decision-maker approaches the 'dispute in a non-partisan frame of mind'.²² It also means that judges should decide cases according to their merit, and judges should not be influenced by factors outside of the law. A key theme

²¹ W Lucy, 'The Possibility of Impartiality' (2005) 25 OJLS 3, 7.

²² *ibid*, 12.

of this thesis is the development of quantitative measurements of the impartiality of the Court of Appeal.

The importance of impartiality in judicial decision-making is such that, no matter how complicated or important the issues, or how preeminent the judge, the appearance or possibility of a lack of impartiality renders any decision void. This is because impartiality is what provides the decisions of judges with authority and legitimacy within the legal and political system. A robust example of this is the series of *Pinochet* cases. *Pinochet* concerned Senator Augusto Pinochet, the former President of Chile. Following a warrant issued in Spain, he was arrested during a visit to London on suspicion of the murder, torture, and hostage-taking of the citizens of Chile during his dictatorship. In *R v Bow Street Metropolitan Stipendiary Magistrate ex parte Pinochet (Pinochet No 1)*,²³ the House of Lords was asked to decide whether Pinochet, as a former Head of State, was entitled to diplomatic immunity for offences as serious as murder, torture, and hostage-taking. The House decided, by a 3:2 majority, that Pinochet was not entitled to diplomatic immunity because, 'it hardly needs saying that torture of his own subjects, or of aliens, would not be regarded by international law as a function of a Head of State'.²⁴

That decision was declared void in *Pinochet (No 2)*,²⁵ when it was discovered that one of the judges in the majority, Lord Hoffman, was a director of Amnesty International which was added as an interested party to the litigation. Holding that Lord Hoffman should have recused himself, Lord Hope said:

'The connections which existed between Lord Hoffmann and Amnesty International were of such a character, in view of their duration and proximity, as to disqualify him on this ground ... he could not be seen to be impartial.'²⁶

²³ [2000] 1 AC 61.

²⁴ *ibid* 108-9.

²⁵ [2000] 1 AC 119.

²⁶ *ibid* 143.

The decision in *Pinochet (No 1)* was set aside on the ground of lack of the appearance of impartiality. In *Pinochet (No 3)*,²⁷ the House concluded that Pinochet could not have immunity for offences said to have been committed after the ratification, in December 1988, of the *International Convention against Torture and other Cruel, Inhuman or Degrading Treatment or Punishment*. The importance of *Pinochet* in relation to impartiality is that it reiterates the absolute requirement of impartiality and the appearance of impartiality, however serious the issues raised by the litigation. It confirms that everybody, however serious the allegations against them, is entitled to have legal action against them determined by an impartial judiciary. The requirement of impartiality is onerous, and the consequences of breach are serious. Accordingly, an allegation of a lack of impartiality is not to be made lightly.

The principle of impartiality will be explored in depth in Chapter 2 of this thesis. The impartiality of the judiciary is a fundamental human right. Article 6 of the *European Convention on Human Rights* (ECHR) provides that:

‘In the determination ... of any criminal charge ... everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law’.

The right to an independent and impartial tribunal is the right which forms the basis of the rule of law and people’s rights before courts. Trechsel and Summers call the right ‘by far the most important guarantee enshrined by Article 6 ... without independent [and impartial] courts there can be no rule of law’.²⁸ The right to an independent and impartial tribunal is so essential that ‘proceedings before a tribunal which does not satisfy the criteria of independence and impartiality can never be fair’.²⁹

²⁷ [2000] 1 AC 147.

²⁸ S Trechsel and S Summers, *Human Rights in Criminal Proceedings*, (Oxford University Press 2006) 46.

²⁹ *ibid*, 47.

In addition to being a substantive right, the impartiality of judicial decision-making relates to what judges do when they make and interpret laws. Cases which reach appeal level are rarely easy cases. There are usually valid legal points to be made on both sides, yet judges are required to reach a decision. What judges do when faced with difficult legal problems is one of the core questions of jurisprudence. Scholars such as Hart and Dworkin, and lawyers such as Bingham, and the Legal Realists, considered the power of judges to decide cases and how it arises. Central to these questions is an analysis of the indeterminacy of law and the presence of judicial discretion. If the law is always determinate, and judges have no discretion, a lack of impartiality is highly improbable as the outcome of cases may flow in a deductive manner from previous law. Legal theory must explain the role of discretion in judicial decision-making. Aspects of legal theory concerned with judicial impartiality are considered in Chapters 2 and 3 of this thesis.

1.3 An analysis of murder and rape appeals against conviction

The Court of Appeal hears appeals against conviction and sentence from Crown Courts. This thesis does not analyse the Court's decision-making for all offences, but includes only murder and rape appeals against conviction. It does not consider appeals against sentence. A total of 472 appeals against conviction, decided between January 2006 and December 2010, are included in the sample. This is all available murder and rape appeals against conviction decided in this period. Table 1.1 compares data provided in the Official Statistics,³⁰ regarding the total number of appeals heard between 2006 and 2010, and the number of murder and rape appeals analysed in this study.

³⁰ Data relating to the overall workload of the Court between 2006 and 2010 taken from MOJ 'Court Statistics (Quarterly) January to March 2014', 'additional tables', table 5.8. Available from www.gov.uk/government/statistics/court-statistics-quarterly-january-to-march-2014 <accessed 29 July 2016>.

Table 1.1: Comparison of overall Court of Appeal workload and murder and rape workload (2006-2010).

	2006	2007	2008	2009	2010	Totals
Sampled Appeals	120	91	66	91	104	472
Overall workload	572	523	438	430	496	2459

Although only murder and rape appeals are analysed in this study, they accounted for 19% of the overall conviction appeal workload of the Court in this period. It is not sought to claim that this is a representative sample of all decision-making in the Court of Appeal, and as such the results of this study are specific to the Court's decision-making in relation to murder and rape appeals in this period. Furthermore, the results should not be extrapolated to cases outside of the Court of Appeal, as there is a selection effect. As is discussed in Chapter 5, the offences of murder and rape were carefully chosen with the intention that the Court's decision-making in relation to these two specific offences can be analysed. In particular, these offences were chosen because of their character as serious and stigmatic offences. The stakes, for the Court of Appeal, wider society, appellants, complainants, and the deceased and their families, are particularly high for these offences. As such, it is appropriate to consider in closer depth the Court's decision-making in relation to these two specific offences.

1.4 Thesis overview

The aim of this thesis is to analyse the decision-making of the England and Wales Court of Appeal (Criminal Division). In particular, it is sought to explore the impartiality of the Court's decision-making in relation to the offences of murder and rape. A selected range of variables offer an indirect measurement of the principle of impartiality. This study seeks to provide a quantitative analysis of the decision-making of the Court in order to make an original contribution to previous work.

Chapter 2 analyses the meaning of impartiality from a number of differing perspectives. Chapter 3 explores the background and theoretical frameworks of this study. Previous, primarily American, studies of judicial decision-making and models of judicial behaviour are analysed in order to show what has driven the collection for variables used in this thesis. Chapter 4 provides a critical analysis of the Court of Appeal, including its history and powers. This includes a comprehensive analysis of the 'unsafety test'. Chapter 5 explains more explicitly the methods employed in conducting this study. This includes principles of data collection and explanations of the statistical analysis employed. Chapter 6 recounts fully how the individual variables were collected. This includes an evaluation of the template which was used to collect the data, and a full recounting of all the independent variables which were collected. Chapter 7 presents the results of the study. This includes analysis of the grounds of appeal raised; analysis of the predictive power of each independent variable; and the binary logistic regression analyses. Chapter 8 analyses the results of the thesis and evaluates the strength of the evidence regarding the impartiality of the Court of Appeal. It also evaluates how well this study has captured the principle of impartiality, and how the limitations of the method frame the conclusions which can be reached. Chapter 9 concludes the thesis.

1.5 Conclusion

This thesis makes an original contribution to knowledge by conducting a quantitative analysis of the decision-making of the England and Wales Court of Appeal (Criminal Division). This study addresses the level of confidence that observers of the Court can have in the impartiality of the Court of Appeal. It does this by determining whether variables collected from each case are associated with the decision to allow or dismiss an appeal against conviction. The strength and direction of any association found in the data are analysed, to determine what conclusions can be drawn regarding the impartiality of the Court of Appeal.

Chapter 2

The Concept of Impartiality

Introduction

This chapter critically analyses the concept of impartiality, which is the normative question addressed in this thesis. The chapter begins by discussing the important political and constitutional role played by the judiciary; that of upholding the rule of law. In upholding the rule of law, however, judges must also act in accordance with it. One important component of the rule of law is impartiality. Impartial decision-making is a central component of the 'legal model' of judicial decision-making. Several related conceptions of impartiality will be explored. Firstly, this chapter offers general definitions of impartiality and what it means for courts to be impartial, and what impartiality means in the context of this thesis. As discussed in this chapter and in Chapter 3, the impartiality of judicial decision-making has been explored in previous empirical legal studies in other jurisdictions, but rarely in Britain. Aspects of the impartiality of the Court of Appeal have been measured in this thesis by collecting data from Court judgments for statistical analysis.

The role of impartiality in legal theory is then analysed, by considering the writing of Lord Bingham, Hart, and Dworkin. These three writers are analysed because they made some important contributions to the analysis of judicial decision-making and impartiality. By highlighting the central role which impartiality plays in legal theory, it will be shown why it is important that the impartiality of judicial decision-making is studied. As well as being a jurisprudential concept, impartiality is a substantive right under Article 6 of the *European Convention on Human Rights*. The European Court of Human Rights has divided impartiality into two elements: subjective and objective. The subjective aspect requires an assessment of whether the judge was actually free of personal bias. Two issues relating to the subjective impartiality aspect are analysed in this chapter: the issue of (a lack of) diversity, and the psychology of decision-making.

The objective component of impartiality within Article 6 asks whether an objective observer would doubt the impartiality of the court in question, if aware of the facts. This demonstrates the importance of appearances. The Court of Appeal (Criminal Division) has held other parts of the criminal justice system, such as the jury, and the trial judge, to the same standards. The objective test of impartiality allows for quantitative analysis because it asks whether there is any reason to doubt impartiality. It is impartiality in this objective sense which is most closely addressed in this thesis.

2.1 The constitutional role of judges

Traditional Diceyan constitutional theory stipulated the absolute sovereignty of Parliament to make and unmake any laws.³¹ A key source of the authority of Parliament is the House of Commons, which gains its authority through its representativeness of the electors.³² Whilst Dicey perceived the sovereignty of Parliament as absolute, he noted that the courts play an important role in the creation of law. He said: 'the adhesion of our judges to precedent ... leads inevitably to the gradual formation by the Courts of fixed rules for decision, which are in effect laws'.³³ He did not think that this undermined the sovereignty of Parliament, however, because 'judicial legislation' was a subordinate form of law.³⁴

In more modern times, it is acknowledged that the sovereignty of Parliament cannot be absolute, and that judges play a particular role. In *R (on the application of Jackson) v Attorney General*,³⁵ the Appellate Committee of the House of Lords was asked to determine whether the *Hunting Act 2004* was a valid Act of Parliament, given that it was passed by the procedure in the *Parliament Act 1949*, which itself was passed by the procedure in the *Parliament Act 1911*. The

³¹ AV Dicey, *Introduction to the Study of the Law of the Constitution* (8th edition, Macmillan 1915) xxxvi.

³² *ibid*, 34-5.

³³ *ibid*, 18.

³⁴ *ibid*.

³⁵ 2005 UKHL 56.

Parliament Acts allow bills to become law without the assent of the Parliamentary House of Lords in certain circumstances.

The House decided unanimously that the 1949 Act was a valid Act of Parliament, and as such so was the Hunting Act, despite the Parliamentary House of Lords having not approved either Act. Lord Steyn observed that on a strictly legalist approach the decision could mean that the government, which will usually have a majority in the House of Commons, could make constitutional changes, such as altering the composition of, or abolishing, the House of Lords.³⁶ Moreover, the Parliament Act 1949 'could also be used to introduce oppressive and wholly undemocratic legislation' which would not require the assent of the Parliamentary House of Lords.³⁷ He opined that if a government was to seek to introduce 'oppressive' legislation, it would be the distinct role of the judiciary to determine whether it was an action which Parliament, despite being sovereign, could not take.³⁸ Accordingly, 'the classic account given by Dicey of the doctrine of the supremacy of Parliament ... can now be seen as out of place in the modern United Kingdom'.³⁹ In particular, it is the judiciary which checks the sovereignty of Parliament.

Other members of the House pointed to a similar role for the judiciary. Lord Hope said that 'the courts have a part to play in defining the limits of Parliament's legislative sovereignty'.⁴⁰ In particular, 'the rule of law enforced by the courts is the ultimate controlling factor on which our constitution is based'.⁴¹ Baroness Hale referred to the courts treating with suspicion any attempt by Parliament to 'subvert the rule of law'.⁴²

³⁶ *ibid*, [101]. See also, Lord Steyn, 'Democracy, the Rule of Law, and the Role of Judges' (2006) EHRLR 243.

³⁷ *ibid*, [102].

³⁸ *ibid*.

³⁹ *ibid*.

⁴⁰ *ibid*, [107].

⁴¹ *ibid*.

⁴² *ibid*, [159].

In other cases the judiciary have expressed their role in enforcing the rule of law. In *A v Secretary of State for the Home Department*⁴³ Lord Bingham said ‘the function of independent judges to interpret and apply the law is universally recognised as a cardinal of a modern democratic state, a cornerstone of the rule of law itself’.⁴⁴ In *R v Horseferry Road Magistrates ex parte Bennett*⁴⁵ the House of Lords accepted that part of the judicial role is to ‘oversee executive action and to refuse to countenance behaviour that threatens either basic human rights or the rule of law’.⁴⁶ Finally, the Court of Appeal in *R v Mullen*⁴⁷ stressed the ‘considerable weight’ attached to the Court’s role in discouraging the disregarding of the rule of law which occurred in that case.⁴⁸

The judiciary’s role in enforcing the rule of law is particularly pertinent to criminal appeals, and particularly pertinent to criminal appeals for the most serious offences. This is because it is in the enforcement of the criminal law that the coercive power of the State is particularly evident. In murder and rape appeals, the stakes are particularly high given the extensive loss of liberty which usually follows if convicted of either of these offences. As is discussed in Chapter 4, the specific role of the Court of Appeal is to seek to correct miscarriages of justice, in the sense of the conviction of the innocent, and to uphold the rule of law.⁴⁹ As the Court has this important role, it is essential that it acts in accordance with it itself. As is now explained, the impartiality of judicial decision-making is a core element of the rule of law.

2.2 The rule of law paradigm and the legal model

Geyh referred to there being a ‘rule of law paradigm’ in relation to the decision-making of judges.⁵⁰ The paradigm ‘features competent, honest, impartial, and

⁴³ 2004 UKHL 56.

⁴⁴ *ibid*, [42].

⁴⁵ [1994] 1 AC 42.

⁴⁶ *ibid*, 62.

⁴⁷ [1999] 2 Cr App R 143.

⁴⁸ *ibid*, 156-7.

⁴⁹ See JR Spencer, ‘Quashing Convictions for Procedural Irregularities’ (2007) Crim LR 835.

⁵⁰ CG Geyh, *Courting Peril: The Political Transformation of the American Judiciary* (Oxford University Press 2016) 2.

independent judges who interpret and apply the rules ... by bracketing out extralegal influences from within and without, and following the law on a case-by-case basis'.⁵¹ The legal model, similarly, posits that 'judges decide cases through the systematic application of external, objective sources of authority'.⁵² These sources of authority are the rules, standards, and principles embodied in statute and case law.⁵³ Judges rely upon 'the canons of constructions, or perhaps a sense of the purpose underlying the statute' to decide how to apply the law.⁵⁴ The legal model is in conflict with behavioural and attitudinal models, which argue, owing to the indeterminacy of law, that judges can and do seek to achieve their policy goals. For behaviouralists and attitudinalists, the legal model serves to '[rationalise] ... Court's decisions and to cloak the reality of the Court's decision-making process'.⁵⁵

As discussed further in Chapter 3, the alternative models to the legal model, such as the behavioural or the attitudinal, developed following the American Legal Realism movement. The American Legal Realists sought to debunk a version of the legal model, which they called formalism or functionalism. During the 'formalist era', said to exist in American law schools between 1870 and 1920, it was thought that judging was mechanical or syllogistic, and legal rules were determinate.⁵⁶ The Realists doubted that legal rules determined legal outcomes.⁵⁷ However, as Leiter discussed, the Legal Realist version of formalism was rather 'vulgar' and few scholars were likely to have subscribed to a view that law operated syllogistically.⁵⁸ As is discussed below, Ronald Dworkin offered a view that the law, which he conceived as including moral principles, was determinate, but he was not a vulgar formalist who viewed law as mechanical.

⁵¹ *ibid*, 16.

⁵² FB Cross, 'Decisionmaking in the US Circuit Courts of Appeals' (2003) *Cali L Rev* 1459, 1462.

⁵³ *ibid*.

⁵⁴ *ibid*, 1463.

⁵⁵ JA Segal and HJ Spaeth, *The Supreme Court and the Attitudinal Model Revisited* (Cambridge University Press 2002) 53.

⁵⁶ See BZ Tamanaha, *Beyond the Formalist / Realist Divide: The Role of Politics in Judging* (Princeton University Press 2009) 1.

⁵⁷ Explored in Chapter 3 of this thesis.

⁵⁸ See B Leiter, 'Legal Formalism and Legal Realism: What is the Issue?' (2010) *LT* 111.

The ELS community has been accused of treating the legal model as meaning the same thing as vulgar formalism. Smith, for instance, argued that Segal and Spaeth's *The Supreme Court and the Attitudinal Model*⁵⁹ used vulgar formalism as a straw man version of the legal model.⁶⁰ Smith argued that Dworkin's version of formalism, which rejects mechanical jurisprudence but claims to limit the amount of decision-making discretion judges have, would be a more robust target against the attitudinal model.⁶¹ The legal model, therefore, is not vulgar formalism but posits that judicial discretion is in some way constrained by legal materials, and maintains that 'politically motivated judicial decisions, unconstrained by precedent or reason, are incompatible with the legal model'.⁶²

A definition of the legal model in the context of UK courts has recently been offered by Arvind and Stirton,⁶³ They argued that:

'Courts typically operate through relatively open-textured concepts, such as "reasonableness" ... Whilst these concepts are not capable of precise definition and hence give the judiciary some flexibility in deploying them in deciding cases, their application in a given case is nevertheless guided by precedent, which constrains the ability of judges to simply decide cases in accordance with their ideological preferences'.⁶⁴

A test of the legal model, therefore, is how constrained the judges are by precedent and other legal rules. The legal model cannot be dismissed by pointing out that judges exercise discretion, and so the law is not determinate. This is because the legal model can still operate when the law is indeterminate if the judges are constrained or guided by existing law, or the intention behind the law.

⁵⁹ Segal and Spaeth (n 55 above).

⁶⁰ See R Smith, 'Symposium: The Supreme Court and the Attitudinal Model' (1994) 4 Law and Courts Newsletter 8. In *The Supreme Court and the Attitudinal Model Revisited* (n 55 above) they do consider the implications of Dworkin and 'postrealist formalists'.

⁶¹ *ibid.*

⁶² FB Cross, 'Political Science and the New Legal Realism: A Case of Unfortunate Interdisciplinary Ignorance' (1998) 92 Nw U L Rev 251, 263.

⁶³ TT Arvind and L Stirton, 'Legal Ideology, Legal Doctrine and the UK's Top Judges' (2016) PL 418.

⁶⁴ *ibid.*, 422.

The Court of Appeal operates the ‘unsafety test’ which, as discussed in Chapter 4, is an ‘open-textured’ concept. A measure is used in this thesis to capture how certain rules operate with the ‘unsafety test’, and if it operates to constrain or guide the operation of the judges’ discretion in the way the legal model would suggest. These measures include how strongly the presence / absence of errors is correlated with outcomes, and how strongly *Criminal Cases Review Commission* (CCRC) appeals are correlated with outcomes. If these measures do constrain judicial discretion, a key question addressed in this thesis is whether there remains room for factual and demographic variables to have an association with the outcome of appeals and what this means for the impartiality of the Court.

According to Geyh, the ‘rule of law paradigm’ is in the process of erosion in the United States, because social science research (especially ELS) has demonstrated a relationship between judicial decision-making and a range of factors, including extra-legal factors.⁶⁵ In the UK, in contrast, there does not appear to be such an erosion of the rule of law paradigm. Thomas argued that this is due to a ‘well-entrenched reluctance to acknowledge that the judiciary has political significance’.⁶⁶ Robertson pointed to the lack of a written constitution in Britain (and the presence of one in the US) as being a reason for the difficulty in exploring the political significance of the British judiciary.⁶⁷ It is easier to categorise a judicial opinion as being in a particular political direction when the judgment is based upon an assessment of a single document which uses political language and confers political rights.⁶⁸

Griffith’s study⁶⁹ of how politics affects the decision-making of the senior British judiciary is one important exception to this rule. Griffith argued that judges’ conception of ‘public interest’ concerns preserving the interests of the State; law

⁶⁵ See Geyh (n 50) at 80, and especially Chapter 3.

⁶⁶ C Thomas, *Judicial Diversity in the United Kingdom and Other Jurisdictions* (Commission for Judicial Appointments 2005) 21.

⁶⁷ D Robertson, ‘Judicial Ideology in the House of Lords: A Jurimetric Analysis’ (1982) 12 *Brit J Pol Sci* 1, 7.

⁶⁸ *ibid.*

⁶⁹ JAG Griffith, *The Politics of the Judiciary* (5th edition, Fontana 1997).

and order; and the judges' views on the social and political issues of the day.⁷⁰ He argued that the senior judiciary cannot be politically neutral because they are required to make political choices.⁷¹ Moreover, owing to the particular senior position which judges hold in society, their interpretation of where the public interest lies is typically conservative.⁷² Griffith's study is not directly comparable to American attitudinal studies which have documented the impact of judicial politics on American judging, leading to the erosion of the 'rule of law paradigm'. This is because his work was not a systematic quantitative study, but was doctrinal in nature and focused upon particular cases which he felt indicated the judiciary's political inclinations. Accordingly, Griffith's publication did not lead to a sustained development of quantitative empirical judicial studies in Britain.⁷³

As there is no sustained culture of judicial studies in Britain, it is unclear how much relevance American studies, and especially the models based upon them, will have to judicial decision-making in the UK. As such, this thesis does not test directly whether any model of judicial decision-making explains the decisions of the Court. It instead explores in a broad manner one central tenant of the rule of law paradigm: impartiality. As there is an absence of evidence to the contrary, it is assumed in this thesis that the Court did decide in accordance with the legal model, and did act impartially. Any allegation of a lack of impartiality is a serious allegation to level against the judiciary. Accordingly, strong evidence would be needed to justify any claim that the Court lacked impartiality. This thesis seeks to explore whether there is sufficient evidence to question the appearance of the Court's impartiality, by utilising variables from a range of models of judicial decision-making.

⁷⁰ *ibid*, 297.

⁷¹ *ibid*, 336.

⁷² *ibid*.

⁷³ See Thomas (n 66) at 21.

2.3 What is impartiality?

Garner's *Dictionary of Legal Usage* defines 'impartial' as being a synonym of 'fair'.⁷⁴ 'Fair' has a number of other synonyms: 'just; equitable; disinterested; dispassionate; objective.'⁷⁵ These adjectives describe judges who 'have no personal stake or bias in an outcome and who apply the proper standards without improper influences'.⁷⁶ Impartiality could be considered as expressing a general concern with fairness and equality between the parties to litigation. This study seeks to provide a partial test of impartiality by exploring whether a range of variables are associated with particular outcomes.

Lucy argued that the requirement of impartiality could have somewhat negative connotations.⁷⁷ Few would consider being 'dispassionate' or 'disinterested' to be positive characteristics in most walks of life. Moreover, judges should not be dispassionate or disinterested about their role of achieving the ends of justice. As Griffith noted, judges 'frequently appear – and speak – as men with weighty, even passionate views of the nature of society and the content of laws', and so do not appear wedded to a 'sanctified' view of impartiality.⁷⁸ Impartiality, therefore, must mean something specific in the decision-making process when it relates to the requirement that judges are impartial.

Lucy argued that in an adjudicative setting, impartiality relates to the judges being impartial as to the outcome, meaning that outcomes are produced regardless of the needs and status of the parties.⁷⁹ Moreover, to be impartial, decisions must 'pay no heed to past or present deeds of disputants', and 'impact to exactly the same degree on disputants'.⁸⁰ In addition to being impartial as to outcomes, in an adjudicative setting procedural impartiality is required.⁸¹ This highlights 'a property of the rules and practices that constitute many decision-making

⁷⁴ BA Garner, *Garner's Dictionary of Legal Usage*, (3rd ed, Oxford University Press 2011) 428.

⁷⁵ *ibid*, 351.

⁷⁶ *ibid*.

⁷⁷ W Lucy, 'The Possibility of Impartiality' (2005) 25 OJLS 3, 14-5.

⁷⁸ JAG Griffith, *The Politics of the Judiciary* (5th edition, Fontana 1997) 284.

⁷⁹ Lucy (n 77) 8.

⁸⁰ *ibid*.

⁸¹ *ibid*, 11.

processes'.⁸² Procedural impartiality 'requires some limitation upon the type of consideration relevant in judicial decision-making',⁸³ in particular, to be impartial, 'adjudicative outcomes are arrived at by reference to the applicable law'.⁸⁴

The requirement that outcomes are arrived at by reference to the applicable law, Lucy argued, 'protects and promotes impartiality because these factors serve to exclude others'.⁸⁵ This means that 'outcomes ought not to be determined by judicial empathy with the values, prejudices, principles and conceits of heterosexual, white upper middle class protestant men'.⁸⁶ Such a conception of impartiality permits some analysis of whether this does accord with the actual decision-making of judges in Courts. As will be explored in Chapters 3, 5 and 6, impartiality has been measured by determining whether certain factors are predictors of the outcome of appeals against conviction.

The judiciary, unlike Parliamentarians in the House of Commons, do not face election and so their legitimacy must emanate elsewhere. The belief that the British judiciary produces high quality and impartial decisions provides it with legitimacy. In Tyler's⁸⁷ 1984 study of the role of legitimacy in ensuring compliance with the law, he claimed that 'those who regard legal authorities as having greater legitimacy are more likely to obey the law in their everyday lives'.⁸⁸ This finding is in contrast with a common assumption at the time that people are motivated to obey the law due to instrumental concerns, such as the threat or the deliverance of sanctions.⁸⁹ Tyler's research since 1984 has continued to suggest that non-instrumental factors cause compliance with the law.⁹⁰ He found that if people perceive processes as being fair they see it as legitimate regardless of what the outcome of the case was.⁹¹ Thus, if a law or authority is seen as being

⁸² *ibid.*

⁸³ *ibid.*, 23.

⁸⁴ *ibid.*

⁸⁵ *ibid.*, 24.

⁸⁶ *ibid.*

⁸⁷ TR Tyler, *Why People Obey the Law*, (2nd ed, Princeton University Press 2006).

⁸⁸ *ibid.*, 57.

⁸⁹ *ibid.*, 21.

⁹⁰ *ibid.*, 269-288.

⁹¹ *ibid.*, 101.

procedurally fair there is greater compliance and acceptance. The perception of impartiality, which was measured by asking respondents whether the officials they dealt with acted improperly or dishonestly, whether they lied, or whether they treated the respondent badly because of their age, sex, race or nationality, showed a link with whether the respondents saw the official as being procedurally just, and so worthy of being obeyed.⁹² As Tyler said, ‘the elements of procedural justice most directly linked to decision-making are judgments about the neutrality of the decision-making procedure.’⁹³

Later research by Casper, Tyler and Fisher showed that those charged with serious crimes consider that the procedural fairness of the proceedings is linked to how satisfied they are with the process.⁹⁴ Research by Sunshine and Tyler showed that people were more likely to obey the police if they viewed them as behaving fairly and as having legitimacy.⁹⁵ A study by Paternoster and his colleagues showed that recidivism rates in spousal assault cases were lower if the accused believed he was treated fairly by the police.⁹⁶ Thus, given the Court of Appeal’s role in reviewing the outcomes of the criminal justice process, it is especially important that the Court fulfils that role with impartiality, if it is to present an image of legitimacy.

Tyler’s research was concerned with how a government can best achieve general compliance with its laws. This is a different focus from this thesis, as the Court of Appeal does not demand general compliance from the public, or even if it does, its reach is minimal. However, Tyler’s research on the importance of legitimacy is relevant because the Court of Appeal sits atop the criminal justice system, which does demand general compliance. The importance of impartiality in relation to legitimacy is that a court lacking in impartiality cannot be legitimate. If a court lacks legitimacy, it does not matter what the court’s decisions are. In this

⁹² *ibid*, 128-130.

⁹³ *ibid*, 137.

⁹⁴ JD Casper, TR Tyler and B Fisher, ‘Procedural Justice in Felony Cases’ (1988) (22) *Law and Soc’y Rev* 483.

⁹⁵ J Sunshine and TR Tyler ‘The Role of Procedural Justice and Legitimacy in Shaping Support for the Police’ (2003) (37) *Law and Soc’y Rev* 513.

⁹⁶ R Paternoster, R Bachman, R Brame and LW Sherman, ‘Do Fair Procedures Matter? The Effect of Procedural Justice on Spouse Assault’ (1997) 31 *Law & Soc’y Rev* 163.

way, the legitimacy of the court is a question which comes before the question of how well the court performs in fulfilling its functions. It is this prior question which is addressed in this thesis: the Court of Appeal's impartiality, and so its legitimacy.

2.4 Judicial discretion and impartiality

In his 2006 Maccabean Lecture,⁹⁷ Lord Bingham said that the traditional view of judging, which he said was only partly true, was that judges should not be motivated by 'extraneous considerations'.⁹⁸ If they were, this would be incompatible with impartial decision-making. These extraneous considerations may be the 'prejudice or predilection of the judge, or, worse, any personal agenda of the judge, whether conservative, liberal, feminist, libertarian, or whatever'.⁹⁹ The traditional view of judging saw judicial decision-making as being narrowly concerned with deducing outcomes from legal logic.¹⁰⁰ This view therefore holds that judges do not make law, but only declare it.¹⁰¹

Lord Bingham argued that modern judges reject this interpretation of the role of the judge. He said most judges acknowledge that 'judges do make law, and regards it as an entirely proper judicial function, provided it is exercised within certain limits'.¹⁰² Most judges are said to subscribe to this understanding of the judicial role because 'the [traditional] approach is radically inconsistent with the subjective experience of judges'.¹⁰³ They know, he said:

'That the cases which come before them do not in the main turn on sections of statutes which are clear and unambiguous ... they know, and the higher the Court the more right they are, that decisions involve issues of policy'.¹⁰⁴

⁹⁷ Available from www.law.cf.ac.uk/newsandevents/transcripts/271005.pdf <accessed 8 November 2015>.

⁹⁸ *ibid*, 28.

⁹⁹ *ibid*.

¹⁰⁰ *ibid*, 44.

¹⁰¹ See T Bingham, *The Business of Judging: Selected Essays and Speeches* (Oxford University Press 2000) 25.

¹⁰² *ibid*, 27.

¹⁰³ *ibid*, 28.

¹⁰⁴ *ibid*, 28.

Thus, according to Bingham, judges accept that in cases which reach court there will be arguments of some strength available to both sides. If there were not arguments available to both sides the litigation would never happen, as one side would concede, or settle, or not commence proceedings. The fact that access to the senior judiciary often requires some kind of permission process, where another judge or body determines whether there is legal merit, further demonstrates that there will usually be arguments on both sides.¹⁰⁵

Lord Bingham argued that judicial discretion appears when the application of the law does not determine the issue, and so the resolution of the case 'depends upon the individual judge's assessment of what is fair and just to do in the individual case'.¹⁰⁶ Judges have no discretion when determining issues of fact or of law, but when having made such findings he or she has to choose between different courses of action, for instance, to allow or to dismiss an appeal against conviction, a discretion is exercised.¹⁰⁷

The exercise or presence of discretion can appear 'slightly deviant'.¹⁰⁸ Discretion is troubling for the criminal justice system because it potentially allows its agents to 'engage in discriminatory activities and to subvert policies they do not agree with'.¹⁰⁹ In the Court of Appeal, this might include, for instance, judges orientated towards crime control quashing fewer convictions; or more liberal judges quashing more convictions. As Wendel says, 'principles of judicial impartiality must take a position on the existence of judicial discretion and the problem of legal interpretation'.¹¹⁰ Both Hart and Dworkin sought to explain how judicial preferences are excluded from the decision-making processes in their

¹⁰⁵ This is seen in the 'leave' process to the Court of Appeal; the referral process of the CCRC; and 'admissibility' process of the European Court of Human Rights, and so on.

¹⁰⁶ Bingham (n 101 above) 36.

¹⁰⁷ *ibid.*

¹⁰⁸ DJ Galligan, *Discretionary Powers: A Legal Study of Official Discretion* (Clarendon 1990) 1.

¹⁰⁹ L Gelsthorpe and N Padfield (eds) *Exercising Discretion: Decision-making in the Criminal Justice System and Beyond* (Willian Publishing 2003) 2.

¹¹⁰ WB Wendel 'Impartiality in Judicial Ethics: A Jurisprudential Analysis' (2012) 22 *Notre Dame LJ Ethics & Pub Pol'y* 305, 307.

theories.¹¹¹ Their theories regarding the nature of law placed some importance on adjudication and how judges decide what the law is, or what it ought to be.

Hart sought to limit the possibility of judicial discretion by arguing that law is a rule governed system, and the outcomes of litigation or disputes will usually be determined by following precedent or statutes.¹¹² However, Hart, like Dworkin, was not a 'positivistic formalist'¹¹³ who believed in mechanical decision-making. He argued that at times judicial decision-making requires the exercise of discretion when cases fall within the 'open-texture' of the rule.¹¹⁴ Hart argued that there are many 'clear cases' where general expressions are clearly applicable, such as a motor-car is clearly a vehicle.¹¹⁵ If a question was raised as to whether a car was permitted to enter a park which stated that there are 'no vehicles allowed', the judge would have no discretion, because a car is clearly a vehicle. However, it would be more difficult to decide whether an electric toy car is also a vehicle under such a rule. If a dispute arose as to whether this was a vehicle, 'we confront the issues at stake and can then settle the question by choosing between the competing interests in the way which best satisfies us'.¹¹⁶ Such a case would be a hard case, and the decision-maker would have to make a choice, or exercise discretion or judgment, and the judge's answer would be legislative in nature, to the extent that once the decision is made there would then be governing precedent constituting a new rule.

According to Hart, this discretion exists only at the limits, or the 'penumbra', of the rule. It is only when the law 'runs out' that it is possible that partiality could occur as it is only here where judges exercise discretion.¹¹⁷ For Hart, this 'penumbra' is small and is a natural consequence of the nature of language and

¹¹¹ See O Raban, *Modern Legal Theory and Judicial Impartiality* (Glasshouse Press 2003) 2.

¹¹² What Hart meant by a 'rule', and how they are identified by the 'rule of recognition', is not relevant to the present discussion.

¹¹³ See GR Carrio 'Professor Dworkin's Views on Legal Positivism' (1980) 55 *Ind L J* 209, 221.

¹¹⁴ HLA Hart, *The Concept of Law*, (3rd ed, Oxford University Press 2012).

¹¹⁵ *ibid*, 129.

¹¹⁶ *ibid*.

¹¹⁷ *ibid*, 130.

the way that society develops. New scenarios may arise which were never considered by Parliament at the time they enacted a rule.

Hart's theory of law as being characterised by rules which are largely determinate can be disputed by challenging the size of the 'penumbra'. Hart could be wrong to say that judicial discretion exists only at the limits of legal rules. It may instead infect the whole of the law. This is the view of some adherents of Critical Legal Studies, such as Duncan Kennedy who argued that many of the decisions (primarily American) judges reach are ideological in nature.¹¹⁸ The Legal Realist position is similar. American Legal Realism is important to this thesis, and evaluated further in Chapter 3.

In *Taking Rights Seriously*, Ronald Dworkin sought to explain how judicial preferences are excluded from the decision-making process by denying that judges exercise strong discretion.¹¹⁹ Instead, even in hard cases, one party will have a right to win, and the judges have a duty to 'discover what the rights of the parties are, not to invent new rights retrospectively'.¹²⁰ Judges do this because it is not only legal rules, but also principles which provide citizens with legal rights. Principles, according to Dworkin, are standards to be observed as a 'requirement of justice, or fairness or some other dimension of morality'.¹²¹ As noted above, Dworkin was in some ways a formalist, as he believed that judges do not exercise strong discretion. However, he did not believe that law was mechanical or rule governed. Rather, the position of the judge was central, and the judges' duty is to discover the answer to a legal question by constructing a theory which best fits and justifies the law.

¹¹⁸ D Kennedy, *A Critique of Adjudication: fin de siècle* (Harvard University Press 1997).

¹¹⁹ R Dworkin, *Taking Rights Seriously* (Duckworth & Co 1977).

¹²⁰ *ibid*, 81.

¹²¹ *ibid*, 22.

To demonstrate this, Dworkin invented judge Hercules: 'a lawyer of superhuman skill, learning, patience and acumen'.¹²² His purpose was to demonstrate how a judge would determine what legal principles require in hard cases. Dworkin said that Hercules would 'construct these theories in the same manner as a philosophical referee would construct the character of a game'.¹²³ In a hard case where rules or principles point in different directions, 'Hercules must turn to the remaining constitutional rules and settled practices under these rules to see which of these two theories provides a smoother fit with the constitutional scheme as a whole'.¹²⁴ Dworkin acknowledged 'many of Hercules's decisions about legal rights depend upon judgments of political theory that might be made differently by different judges or the public at large'.¹²⁵ Unlike Hart, who argued that law can run out due to the open-texture of rules, Dworkin believed that principles also formed part of the law, and these cannot run out.

Dworkin argued that there is no room for partiality in his theory because the determination of which principles apply is not the judge's own personal convictions, and must not be based upon policy, but must be based upon his or her attempt to determine what the law requires. This cannot be attributed to Hercules's personal convictions, but to his attempt to determine what the community's constitutional and political morality is and its fit with the legal rules.¹²⁶ Decision-making is therefore not based on policy, but on judges determining principles. In *Law's Empire*,¹²⁷ Dworkin explains his proposition of 'law as integrity'. Dworkin argued that a judge determines a hard case by deciding which outcome 'follows from the best interpretation of what judges characteristically do about statutes'.¹²⁸ According to Dworkin, there will be one right answer to this question.

¹²² *ibid*, 105.

¹²³ *ibid*.

¹²⁴ *ibid*, 106.

¹²⁵ *ibid*, 123.

¹²⁶ *ibid*, 126

¹²⁷ R Dworkin, *Law's Empire* (Harvard University Press 1986) 253.

¹²⁸ *ibid*, 87.

Law as integrity requires a judge deciding a hard case to think of himself as an 'author in the chain of common-law'.¹²⁹ Judges in earlier cases may not have answered the exact same question as him, but earlier judges will have left precedent which orbit the question in his case. He must then use his own judgment to make the story (the law) as good as it can be as a cohesive whole.¹³⁰ Since the judge is required to interpret the law in the manner of a chain novelist, eventually the wealth of legal materials will become so great that there is only one right legal answer available.

The proposition that there is one right answer to every novel legal question is attractive because it means that partiality is impossible. Dworkin explains that the judge trying to reach the right answer which fits and presents law in its best light, will need to draw upon his or her own conceptions of justice and political morality.¹³¹ If these conceptions are 'right' then he or she will reach the right answer as to what decision should be reached. The difficulty is that Dworkin permits judges to use their own conceptions of justice and morality if it produces the 'correct' answer. Indeed, the decision-making process is a political exercise, with which other judges or citizens may disagree.¹³² This does not appear satisfactory, because, as Raban stated, 'who cares that these personal opinions are somehow "correct" if we all hold contradictory views?'¹³³

For Dworkin, the question is not whether the right answer is the result of the personal opinions of judges, as Dworkin accepts they are, but whether that opinion is right.¹³⁴ But, as Posner argued, 'interesting' moral claims beyond tautological ones such as 'killing is wrong' are always local, in that they are 'relative to the moral code of the particular culture'.¹³⁵ There may therefore be no

¹²⁹ *ibid*, 238-9.

¹³⁰ *ibid*, 239.

¹³¹ Dworkin, *Law's Empire* (Harvard University Press 1986) 239. See also Dworkin *Taking Rights Seriously* (Duckworth & Co, 1977) 127.

¹³² Dworkin, *Taking Rights Seriously* (n 131) 126-8.

¹³³ See O Raban, *Modern Legal Theory and Judicial Impartiality* (Glasshouse Press 2003) 82.

¹³⁴ *ibid*, 80.

¹³⁵ RA Posner, 'The Problematics of Moral and Legal Theory' (1998) (111) (7) *Harv L Rev* 1637, 1642.

'correct' conceptions of political morality. If there are no correct answers, Dworkin's thesis says little about the desirability of impartial decision-making but appears to invite judges to decide cases based on their personal predilections.

Hart's and Dworkin's theories concern the extent of legislative discretion available to judges in hard cases. It may be observed that Lord Bingham's opinion on how judges decide cases may be compatible with Hart's, in that he acknowledges hard judicial discretion. Dworkin sought to argue that judges do not exercise discretion in this sense, as the law (broadly defined to also include principles) was determinate. It is clear, however, that Dworkin in no way proposed a theory of legal formalism in which it is legal *rules* which were determinate. As will be discussed in Chapter 3, the American Legal Realists also explored the level of discretion available to judges. In Chapter 4, the 'unsafety test' will be analysed in order to explore the level of discretion judges in the Court of Appeal have. The 'unsafety test' appears a particularly open-textured concept, which invites discretion. However, closer analysis shows that the discretion in the operation of the test is constrained by rules. In Chapter 8 it is discussed whether any of these theories of judicial decision-making appear to be well reflected in the sample of appeals studied in this thesis.

2.5 Independence and impartiality

Article 6 of the *European Convention on Human Rights* (ECHR) provides that:

'In the determination ... of any criminal charge ... everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law'.

Impartiality appears to mean something separate to independence. Independence relates specifically to the relationship between courts and the executive, whilst impartiality relates to how decisions are reached. In *Findlay v United Kingdom*,¹³⁶ Mr Findlay alleged in the European Court of Human Rights

¹³⁶ *Findlay v United Kingdom* (1997) 24 EHRR 221.

that his Article 6 rights were breached because his trial by court-martial was not an independent or an impartial tribunal.¹³⁷ In considering whether the court-martial was 'independent' the Court said that regard must be had, *inter alia*, to the manner of the appointment of its members and their terms of office; the existence of guarantees against outside pressures and the appearance of independence.¹³⁸ This has its origins from an earlier case decided by the European Court of Human Rights, *Le Compte v Belgium*.¹³⁹ In that case the Court said that the term 'tribunal' is only warranted if the body satisfies the requirement of independence which requires, in particular, independence from the executive and independence from the parties to the case.¹⁴⁰ The Court in *Findlay* explained that there are two aspects to impartiality for the purposes of Article 6. There is a subjective and an objective aspect. The subjective aspect is that the judges of the court should in fact be free of personal prejudice or biases. The objective aspect is that there must be sufficient guarantees from an objective viewpoint 'sufficient to exclude any legitimate doubt as to a lack of impartiality'.¹⁴¹

In *Findlay* the Court found that due to the position of a senior member of the court-martial – the 'convening officer' – it lacked independence and impartiality. The convening officer was the most senior officer involved in the case and it was his role to select the other members of the court-martial, and confirm the result, and he had a role in the prosecution of the offence. Accordingly, 'Mr Findlay's misgivings about the independence and impartiality of the tribunal ... were objectively justified'.¹⁴² Since the right to an independent and impartial tribunal was breached it was not necessary to consider the other claims under Article 6. By stating that the misgivings were objectively justified the Court was not stating that the court-martial lacked independence or that the members lacked actual impartiality. It meant that from an objective viewpoint there were legitimate grounds to question the Court's impartiality.

¹³⁷ *ibid*, [68].

¹³⁸ *ibid*, [73].

¹³⁹ *Le Compte v Belgium* (1982) 4 EHRR 1.

¹⁴⁰ *ibid*, [55]

¹⁴¹ *Findlay*, (n 136) at [73]

¹⁴² *ibid*, [80].

It may be challenged whether there is any difference between the concepts of independence and impartiality. Trechsel and Summers are inclined to think that there is little difference between the two.¹⁴³ Independence and impartiality are related because each depends upon the other. The independence of courts provides the conditions for judges to decide cases impartially. Independence ‘concerns the relationship of the decision maker to the government’.¹⁴⁴ Impartiality ‘refers to what goes on, and appears to go on, in the mind of the decision maker’.¹⁴⁵ In relation to the Article 6 requirement of independence in the sense of the constitutional position of the Court of Appeal, little else will be said in this thesis. There is no challenge to the institutional independence of the Court of Appeal. This thesis is concerned with impartiality in the sense of ‘what goes on in the mind of the decision maker’.

2.6.1 The problematics of subjective impartiality

In *Findlay*, referenced above, the European Court of Human Rights referred to there being two aspects of impartiality: a subjective aspect and an objective aspect.¹⁴⁶ The court facing challenge would need to pass both aspects. The subjective aspect of impartiality means that the judges themselves must in fact be free of personal bias. In *Incal v Turkey*¹⁴⁷ the European Court of Human Rights found that Mr Incal’s trial before a National Security Court breached his Article 6 rights as it was not independent or impartial. In concluding that Article 6 was breached the Court said ‘the [subjective aspect] consists in trying to determine the personal conviction of a particular judge in a given case’.¹⁴⁸ The Court did not consider whether the judge in *Incal* was subjectively impartial, but instead focused upon the objective impartiality test.¹⁴⁹ The Court found that there was objective justification for Incal’s misgivings that the Court lacked impartiality.

¹⁴³ S Trechsel and S Summers, *Human Rights in Criminal Proceedings*, (Oxford University Press 2006) 49.

¹⁴⁴ M Kirby, ‘Judicial Recusal: Differentiating Judicial Impartiality and Judicial Independence’ (2015) 4 *Brit J of American LS*, 1, 1.

¹⁴⁵ *ibid.*

¹⁴⁶ *Findlay v United Kingdom* (1997) 24 EHRR 221 [73].

¹⁴⁷ *Incal v Turkey* (22678/93, 09/06/1998, Grand Chamber).

¹⁴⁸ *ibid.*, [65].

¹⁴⁹ *ibid.*

The principle that judges are subjectively impartial is problematic. As a collective, the issue of a lack of diversity raises concerns about whether judges can be impartial. At the individual level, it is problematic to suppose that any individual can be truly impartial, as, like all other human beings, judges have personal and political opinions which may determine how they view the world. Psychological research has suggested that all human decision-making is subject to bias and heuristics.¹⁵⁰ This research has been applied to judges and demonstrates that judges do not appear to be immune to them.¹⁵¹ The problematics of a lack of diversity, and the question of the psychology of decision-making, are analysed in the following sections.

2.6.2 The problematic of diversity

As Thomas says, lack of diversity raises issues of legitimacy and impartiality.¹⁵² There are three potential threats to legitimacy which could be caused by the lack of diversity within the judiciary. Firstly, the homogenous character of the judges could be a symptom of a lack of equal opportunities for lawyers to be elevated to the judiciary. This focuses upon the difficulty faced by individuals to progress. Secondly, the lack of diversity may damage the perceived democratic legitimacy of the judiciary because it is not reflective of the demographic make-up of society. Thirdly, the lack of diversity may mean that there is a 'different perspective' missing from the judiciary.¹⁵³

It is undeniable that the higher judiciary (in particular) is not diverse. As of April 2015, 19% (21 out of 106) of High Court judges were female.¹⁵⁴ 21% (8 out of 38) Court of Appeal judges were female.¹⁵⁵ There has only ever been one female

¹⁵⁰ D Kahneman, *Thinking, Fast and Slow*, (Penguin 2011). See also A Banforado, *Unfair: The New Science of Criminal Injustice* (Crown Publishers 2015).

¹⁵¹ C Guthrie, JJ Rachlinski and AJ Wistrich, 'Inside the Judicial Mind' (2001) (86) *Cornell L Rev* 777.

¹⁵² C Thomas, *Judicial Diversity in the United Kingdom and Other Jurisdictions: A Review of Research, Policies and Practices* (Commission for Judicial Appointments 2005) 62-3.

¹⁵³ *ibid.*

¹⁵⁴ Data published by the Ministry of Justice, available from < <https://www.judiciary.gov.uk/about-the-judiciary/who-are-the-judiciary/diversity/judicial-diversity-statistics-2015/> <accessed 6 September 2016>.

¹⁵⁵ *ibid.*

judge of the Court of the House of Lords / UK Supreme Court. Overall, as of April 2015, 25% of all judges were female, but the majority of these posts were in the relatively junior judiciary – recorders, circuit judges and district judges.¹⁵⁶ 159 out of 2686 (5.9%) of judges who declared their ethnicity declared it to be black or ethnic minority; three of those were High Court judges and the rest were of lower ranks.¹⁵⁷ There was only one High Court judge whose former profession was not a barrister.¹⁵⁸ Judges from non-barrister professions were more likely to be appointed as district judges, but promotion to the Court of Appeal appears the preserve of former High Court judges. None of the current Court of Appeal judges were previously district judges.¹⁵⁹ This suggests that the primary route to the senior judiciary, via the High Court, remains the preserve of white, disproportionately male, former barristers.

As discussed above, the traditional view is that when a judge, or judges, preside over a case they ought to ignore their personal opinions and seek to decide according to the law. On this view, it should not matter in relation to outcomes or the direction of the law that the judiciary is lacking in diversity. This is because the different way that different judges see the world, if there is a difference, should be forgotten when they come to decide a case and the judges should seek to apply the law. Rackley rejected the notion that judges can forget or ignore their personal opinions when deciding cases.¹⁶⁰ This is because:

‘There are times when all judges must do more than simply follow the rules, when the rules run out ... or at least call for the exercise of some discretion ... in all such cases ... the judge must turn to his or her sense of justice, to an understanding of the judicial function and the purpose of law and the like for a solution’.¹⁶¹

¹⁵⁶ Judicial Office Statistics Bulletin, *Judicial Diversity Statistics*, (Judicial Office, 2015) figure 1.

¹⁵⁷ *ibid*, figure 4.

¹⁵⁸ *ibid*, figure 7.

¹⁵⁹ See judicial biographies <<https://www.judiciary.gov.uk/you-and-the-judiciary/going-to-court/court-of-appeal-home/coa-biogs/>> <accessed 20 November 2015>.

¹⁶⁰ E Rackley, *Women, Judging and the Judiciary: From Difference to Diversity* (Routledge 2013) 130.

¹⁶¹ *ibid*, 131.

Since at times judges can no longer follow rules and must turn to her or his sense of justice, it follows that '*who the judge is matters*'.¹⁶² This is because different judges may have different personalities and philosophies and so will see the world differently. Since the law has 'run out' they cannot be entirely impartial as they need some basis upon which to reach a decision. She argued that 'the different lives and experiences of women and men ... will lead to differences in the attitudes, values and perspectives'.¹⁶³ Since they cannot ignore these values and be impartial, these different perspectives will lead to different understanding and interpretation of the law.¹⁶⁴

The 2013/14 Crime Survey for England and Wales included analysis of the levels of confidence respondents held in the criminal justice system.¹⁶⁵ 'Confidence' was measured by confidence in effectiveness and confidence in fairness.¹⁶⁶ 64% of respondents indicated that they thought the system was fair, whilst only 48% thought it was effective.¹⁶⁷ The Survey question most approximate with impartiality was '*the criminal justice system discriminates against particular groups or individuals*' – 37% of respondents agreed.¹⁶⁸ This may be considered a relatively high percentage of respondents believing that the system was discriminatory.

This finding is potentially corroborated by the European Social Survey which placed the United Kingdom approximately mid-table in the question of how often respondents believed that 'the courts make fair, impartial decisions based upon available evidence'.¹⁶⁹ The UK scored higher than countries such as France, Spain, and the Czech Republic, but not as well as countries such as Germany,

¹⁶² *ibid*, 132. (Emphasis original).

¹⁶³ *ibid* 148.

¹⁶⁴ *ibid*.

¹⁶⁵ See K Jansson, *Public Confidence in the Criminal Justice System: Findings from the Crime Survey for England and Wales 2013/14* (Ministry of Justice 2015).

¹⁶⁶ *ibid*, 2.

¹⁶⁷ *ibid*, 3.

¹⁶⁸ *ibid*, 4.

¹⁶⁹ J Jackson et al, *European Social Survey: Trust in Justice Topline Results from Round 5 of the European Social Survey* (European Social Survey 2011) 9.

Sweden, and Switzerland. It is not known whether the lack of diversity in the British judiciary was what led to the lower ranking of the British courts, but it is one potentially explanatory factor.

A study of user experience of several British tribunals conducted by Genn, Lever, and Gray on behalf of what was the Department for Constitutional Affairs posited a relationship between the composition of tribunals and perceptions of fairness.¹⁷⁰ They found that users from ethnic minorities were less likely to perceive unfairness when the tribunal was composed of mixed ethnicities.¹⁷¹ White tribunal users, in contrast, were slightly more likely to perceive unfairness when the tribunal was of mixed ethnicity.¹⁷² This may suggest that the key to understanding the problematic of diversity in relation to democratic legitimacy is that users of tribunals / courts want the bench to reflect themselves. This is a problem for the higher judiciary in Britain, because of its distinct lack of reflection of broader society.

It may be argued that the lack of diversity in the judiciary means that judges do not decide cases in a truly impartial manner, but in a particular (white, male, etc.) version of impartiality. The *Feminist Judgments Project*,¹⁷³ in which academics wrote alternative feminist judgments, suggested that female judges, or *feminist* female judges,¹⁷⁴ could decide cases differently. The Project wrote additional judgments for 23 cases decided by the House of Lords in a wide range of administrative areas. Nine of the judgments dissented on the outcome, and two of these would have turned a 3:2 split into a 3:3 draw.¹⁷⁵ The remaining judgments would have reached the same conclusions but for different reasons. The 'judgments illustrate powerfully how, even at the same time and within the same

¹⁷⁰ H Genn, B Lever and L Gray, *Tribunals for Diverse Users: DCA Research Series No1/2006* (DCA 2006) 226.

¹⁷¹ *ibid.*

¹⁷² *ibid.*

¹⁷³ C McGlynn, R Hunter and E Rackley (eds) *Feminist Judgments From Theory to Practice* (Hart Publishing 2010).

¹⁷⁴ *ibid.*, 6. (Emphasis original).

¹⁷⁵ *ibid.*, 10.

Court, cases could have been reasoned and decided differently'.¹⁷⁶ This might suggest that the lack of diversity in the judiciary means that as a collective it may not reflect the views of wider society.

Whether female judges do decide cases differently is an empirical question. The evidence is mixed as to whether female judges do decide cases differently to male judges. One example of this occurring is in the Supreme Court case of *Radmacher v Granatino*.¹⁷⁷ In this case, concerning the weight which should be given to prenuptial agreements, Lady Hale noted the gender issue in the case and so wrote a partly dissenting separate judgment.¹⁷⁸ In the United Kingdom there has been little empirical research into the relationship between gender and judicial decision-making. In the United States, research on this issue is far more common. In Boyd, Epstein and Martin's analysis of US courts, they found 30 systematic analyses of the association between gender and judicial decision-making.¹⁷⁹ Of these, one third revealed an association between gender and decision-making; one third showed mixed results; and one third showed no association.¹⁸⁰ In their own study they did find evidence of a 'gender effect' in relation to cases concerning sex discrimination. When a female judge sat alone in a sex discrimination case, *and* where a female sat with male judges, persons claiming sex discrimination were more likely to be successful.¹⁸¹

There are difficulties with using the gender of a judge as the sole factor which could influence how he or she decides cases. Firstly, it reduces everything about that judge to her or his gender. There is a myriad of other potential behavioural factors which can influence the way that a judge sees the world – such as ethnicity; experience; sexuality and parenthood. It may also be questioned whether female judges are that different from male judges. A female judge is not

¹⁷⁶ *ibid.*

¹⁷⁷ [2010] UKSC 42.

¹⁷⁸ *ibid.*, [137].

¹⁷⁹ CL Boyd, L Epstein and AD Martin, 'Untangling the Causal Effects of Sex on Judging' (2010) (54)(2) *American J of Pol Sci* 389.

¹⁸⁰ *ibid.*, 392.

¹⁸¹ *ibid.*

an 'ordinary' female, but she is an elite female and a member of an executive collective of wealthy senior lawyers in the same way that a male judge is an elite male. She is also schooled and socialised in the operation of the law in a similar way that male judges are. As Malleson noted, a female judge may be likely to have more in common with her fellow male judges than the females appearing before her in court.¹⁸² In Rackley's opinion, the search for definitive evidence that female judges decide cases differently due *only* to their gender giving them a 'different voice'¹⁸³ is likely to fail.¹⁸⁴ However, 'it may be one reason why judges sometimes differ'.¹⁸⁵

In this thesis, there is an examination of whether gender is associated with decision-making. This includes an assessment of whether the inclusion of a female judge on the bench is associated with particular outcomes, and whether female victims / complainants are associated with particular outcomes in the Court of Appeal. It is not the only variable considered, however, but it just one aspect of decision-making which is analysed.

2.6.3 Psychology and impartiality

Research conducted by Daniel Kahneman suggests that all kinds of decision-making, not only judicial, involves two hypothetical sources of consciousness, which he labels System 1 and System 2. System 1 tasks are:

'Typically fast, automatic, effortless, associative, implicit (not available to introspection), and often emotionally charged; they are also governed by habit and are therefore difficult to control or modify'.

¹⁸² See K Malleson, 'Justifying Gender Equality on the Bench: Why Difference Won't Do' (2003) (11)(1) Fem LS 1.

¹⁸³ This phrase taken from the seminal work of Carol Gilligan who conceived of the 'different voice' thesis. See C Gilligan, *In a Different Voice: Psychological Theory and Women's Development*, (Harvard University Press 1982).

¹⁸⁴ Rackley (n 160 above) 145.

¹⁸⁵ *ibid.*

Whereas System 2 tasks are:

‘Slower, serial, effortful, more likely to be consciously monitored and deliberately controlled; they are also relatively flexible and potentially rule governed’.¹⁸⁶

Kahneman suggested that a number of phenomenon occur, which he called ‘heuristics’, when people make decisions; and these can lead to fallacious decision-making. For instance, Kahneman says that when people are searching for a solution to a difficult question, ‘system 1 will find a related question that is easier and will answer it’.¹⁸⁷ He calls this ‘substitution’. It is problematic because everybody has ‘intuitive feelings and opinions about almost everything’.¹⁸⁸ System 1 will answer the related question intuitively and System 2 will usually endorse it, because, despite being more deliberative, it is also lazy and the initial answer will appear intuitively attractive.¹⁸⁹ The implications of substitution to judicial decision-making are clear. Judicial decision-making tends to require the resolution of difficult questions. If Kahneman is right, the psychology of decision-making means that judging is not a deliberative process but an intuitive process.

Behavioural research on the mental processes involved in judicial decision-making has been conducted. Guthrie and his colleagues conducted research on five of Kahneman’s heuristics to see if they applied to judges.¹⁹⁰ They found that ‘each of the cognitive illusions influenced the decision-making processes of the judges in our study’.¹⁹¹ This may cause doubt as to whether judicial decision-making can ever be value free or impartial. Benforado concluded that in light of the psychological evidence, it is clear that ‘judges are susceptible to numerous unappreciated biases that influence their perceptions of seemingly neutral

¹⁸⁶ D Kahneman, *Thinking, Fast and Slow* (Penguin 2011) 90.

¹⁸⁷ *ibid*, 97.

¹⁸⁸ *ibid*.

¹⁸⁹ *ibid*, 99.

¹⁹⁰ C Guthrie, JJ Rachlinski and AJ Wistrich, ‘Inside the Judicial Mind’ (2000) (86) *Cornell L Rev* 777.

¹⁹¹ *Ibid*, 780.

facts'.¹⁹² Guthrie and his colleagues conducted a further study, utilising case hypotheticals, of whether judges are able to disregard inadmissible evidence when they reach decisions.¹⁹³ They found that, although performing well in some situations, judges were unable to ignore inadmissible evidence or information.¹⁹⁴ If this is true, it indicates the difficulty in judges being expected to ignore what they know about the outside world when they sit as judges.

Although the findings relating to the potential influence of biases on judging are striking, it is important to note an obvious difficulty. The claim that psychological research 'proves' that impartial decision-making is impossible could itself be, on its own terms, subject to the same biases. It suggests that all decision-making may be irrational, but the researchers making such findings are rational. It may be that the conclusions that judges are overcome by confirmation bias was caused by the researchers being overcome by their own confirmation biases. Mitchell argued that studies applying psychological research to judicial decision-making were unduly pessimistic of the possibility for rational judicial decision-making.¹⁹⁵ It is therefore important to have an element of scepticism when considering such findings.

2.7 Objective impartiality

The principle of objective impartiality has at its stem Lord Hewart's seminal dictum that 'justice should not only be done, but should manifestly and undoubtedly be seen to be done'.¹⁹⁶ This concerns the appearance of impartiality, which is of paramount importance to courts. In *Pullar v United Kingdom*,¹⁹⁷ an elected Scottish local councillor was convicted of offering to take a bribe from a firm of architects and a firm of quantity surveyors, in return for Pullar's influence in the

¹⁹² A Benforado, *Unfair: The New Science of Criminal Justice* (Crown Publishers 2015) 161.

¹⁹³ AJ Wistrich, C Guthrie and JJ Rachlinkski, 'Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding' (2005) 4 U Penn L Rev 1251.

¹⁹⁴ *ibid.*

¹⁹⁵ G Mitchell, 'Taking Behavioralism too Seriously? The Unwarranted Pessimism of the New Behavioral Analysis of Law' (2002) 43 Wm & Mary L Rev 1907.

¹⁹⁶ *R v Sussex Justices ex p. McCarthy* [1924] 1 KB 256.

¹⁹⁷ (1996) 22 EHRR 391.

planning department of the local council.¹⁹⁸ One of the 15 jurors in the trial was a junior staff member of the firm of architects. The European Court of Human Rights found that the trial was subjectively and objectively impartial, and so Pullar's claim failed. The Court accepted, no doubt because of some of the reasons discussed above, that 'it may be difficult to procure evidence with which to rebut the presumption' that the tribunal was free of personal, subjective, bias.¹⁹⁹ Accordingly it was the second, objective, requirement which 'provides a further important guarantee'.²⁰⁰ The way that the Court tested whether the Court was objectively impartial was to determine whether the selection of the juror would have caused an 'objective observer' to doubt the impartiality of the tribunal.²⁰¹ The Court found that it was not clear that an objective observer would doubt the impartiality of the jury, and so the jury was considered impartial.

The 'objective observer' test was further considered in *Gregory v United Kingdom*.²⁰² Gregory, a black man, was convicted by a majority verdict of robbery. While the jury was deliberating the judge received a note stating that the jury was showing racial overtones. The judge gave a direction to the jury that they should put aside any prejudice. Gregory complained to the European Court of Human Rights that his Article 6 rights had been breached due to a lack of impartiality, and argued that the judge should have discharged the jury. The Court held that it must consider whether there were 'sufficient guarantees to exclude any objectively justified or legitimate doubts as to the impartiality of the jury'.²⁰³ The Court found that the judge was not required to do more than he did to dispel Gregory's fears.²⁰⁴ The essential point is that it is not sufficient to simply claim that there are doubts about the impartiality of a tribunal, there must be some further objective basis for doubting the impartiality of the court or tribunal in question.

¹⁹⁸ *ibid*, [7].

¹⁹⁹ *ibid*, [32].

²⁰⁰ *ibid*.

²⁰¹ *ibid* [33].

²⁰² (1998) 25 EHRR 577.

²⁰³ *ibid*, [45].

²⁰⁴ *ibid*, [48].

The principle of the objective, or reasonable, observer is entrenched in domestic law. This is the standard against which the impartiality of the Court of Appeal has been measured in this thesis. Lord Goff in *R v Gough*,²⁰⁵ said that the Court should ask itself whether ‘there was a real danger of bias ... in the sense that [the judge] might unfairly regard ... with favour, or disfavour, the case of a party to the issue under consideration by him’.²⁰⁶ The requirement of a ‘real danger’ of bias, rather than a suspicion of bias, might suggest a slightly higher standard before an institution could be considered to lack objective impartiality. Lord Hope in *Porter v Magill*²⁰⁷ submitted a ‘modest adjustment’ to the test to move it closer to the test employed by most of the Commonwealth, Scotland and Strasbourg.²⁰⁸ Lord Hope’s formulation of the test was whether ‘a fair-minded and informed observer, having considered the facts, would conclude that there was a real possibility that the tribunal was biased’.²⁰⁹

The Court of Appeal has held other parts of the criminal justice system to this standard. In *R v Khan and Others*²¹⁰ the issue in the several conjoined appeals and applications was whether the juries appeared to have lacked impartiality due to the occupation of members of the juries. The appellants appealed their convictions due to the presence of police officers, members of the Crown Prosecution Service, and prison officers in their respective trials. The Court said that if it was established that a juror was partial, or had the appearance of being partial, the convictions must be quashed.²¹¹ The Court dismissed all the appeals, finding that none of the jury members’ employment status meant they appeared partial towards any particular witnesses.

Following this dismissal, the appellants *Hanif* and *Khan* appealed to the European Court of Human Rights.²¹² The applicants accepted that the presence of police

²⁰⁵ [1993] AC 646.

²⁰⁶ *ibid*, 670.

²⁰⁷ [2002] 2 AC 357.

²⁰⁸ *ibid*, [103].

²⁰⁹ *ibid*.

²¹⁰ [2008] EWCA Crim 531.

²¹¹ *ibid*, [8].

²¹² *Hanif and Khan v United Kingdom* (2012) 55 EHRR 16.

officers on the jury was not automatically a breach of Article 6, but they argued on the particular facts that Article 6 was breached.²¹³ This was because police officers from the same force as the police jurors were called as witnesses and their evidence was important to the prosecution case.²¹⁴ The European Court of Human Rights agreed, and found that Article 6 had been breached.²¹⁵ This suggests that the test for objective impartiality is broad, and broader perhaps than the Court of Appeal would like it to be. The European Court of Human Rights stressed the importance that courts inspire confidence in the public and one essential element of that confidence is impartiality.²¹⁶

The test employed by the Court of Appeal when considering an allegation of bias is whether a fair-minded or objective observer would consider there was a possibility or risk of bias (lack of impartiality) after considering the facts – the test in *Porter v Magill*. In *R v Ahmed*²¹⁷ the appeals were dismissed because on an objective appraisal there was no material to give rise to legitimate fears that the jury lacked impartiality.²¹⁸ It also applies to the conduct of the judge who must ‘present the defendant’s case fairly and act in an even-handed and impartial manner’.²¹⁹ In *R v Cordingley*²²⁰ the judge’s behaviour towards the appellant was described as ‘brutal’ by the Court of Appeal.²²¹ The result was that the appellant felt the judge was prejudiced against him; he was not fairly tried and there was a failure of due process.²²²

This thesis analyses the Court’s impartiality by determining whether particular selected variables are predictors of successful or unsuccessful appeals. As stated in Chapter 1, this thesis addresses the question of the impartiality of the

²¹³ *ibid*, [130].

²¹⁴ *ibid*.

²¹⁵ *ibid*, [149].

²¹⁶ *ibid*, [138].

²¹⁷ [2014] EWCA Crim 619. See also *R v Hanrahan* [2015] EWCA Crim 1653.

²¹⁸ *Ahmed*, *ibid*, [35].

²¹⁹ *R v McDonagh* [2012] EWCA Crim 2013.

²²⁰ [2007] EWCA Crim 2174.

²²¹ *ibid*, [13].

²²² *ibid*, [15].

Court in an incomplete manner. The study design is based upon the exploration of correlations between variables and outcomes. Any variables which are shown to be statistically significant predictors of outcomes must be considered carefully to determine what that means about the decision-making of the Court, given the research design. The emergence of any such patterns in the data does not mean that the Court lacked impartiality, but will provide a quantitative basis for analysis of the impartiality of the Court's decision-making. Moreover, it must be asked whether the weight of the evidence presented in this thesis, in light of the strengths and limitations of the methods employed, is sufficient to lead an informed observer to doubt the impartiality of the Court. In order to address this question, the methods themselves must be carefully scrutinised.

2.8 Conclusion

It has been shown that impartiality is a concept which permeates legal scholarship. It is a central tenant in jurisprudence and legal theory; human rights; diversity and legitimacy. This study is an analysis of a working Court of Appeal, which delivers judgments in real cases. The decisions of the Court have substantial implications for the lives of appellants, applicants, complainants and victims and their families. It can also have a significant influence of the development of the law of criminal evidence and procedure. Thus, how the Court performs its functions is an important question. The Court has a role in reviewing and upholding the moral authority of convictions, and giving assent, or otherwise, to the continued punishment of the State's citizens. This is an important role for the Court to play, in which the legitimacy of the Court confers legitimacy on the rest of the system. Impartiality is an important part of the Court having legitimacy. But, is the belief in the legitimacy and impartiality of the Court well placed? It is this question which is assessed in this thesis.

It has been shown that the principle that courts should resolve litigation in an impartial manner is problematic one. It is problematic because judges form a largely homogenous group, and so it is possible that their decision-making is not truly neutral, but is a particular, elite white, male, version of neutrality. Furthermore, judges are likely to be subject to the same internal biases as other

people, meaning that their judgments may not be value-free. On the basis of traditional jurisprudence, judges are expected to forget their previous life experience and resolve disputes in a disinterested way. It has been argued that modern legal theory rejects that judges can decide appeals in this way. This is seen in the Article 6 jurisprudence which places an emphasis upon the appearance of impartiality.

Whether judges decide cases impartially could be seen as an empirical question. In some jurisdictions, in particular the United States, this empirical question has been frequently addressed. In the United Kingdom, it is far rarer that the impartiality of judges is questioned, and then tested in a quantitative way. The next chapter outlines the theoretical frameworks and background principles of conducting quantitative empirical legal research on the decision-making of judges.

Chapter 3

Mapping the Empirical Legal Studies Field

Introduction

This study is situated within the quantitative Empirical Legal Studies (ELS) field of legal studies. The particular branch of ELS with which this thesis is concerned is 'judicial studies', which focuses upon the 'theoretical and empirical study of the choices judges make'.²²³ This thesis is described as a 'positivistic' study of judicial decision-making. This means that the study has been conducted with a view to being systematic and objective in the data collection exercise and analysis. The positivistic study of the decision-making of judges has its roots in the American Legal Realism movement. The Realists challenged the traditional functional / formalism view of how judges decide cases, by arguing that the law had greater indeterminacy than the traditional view could accommodate. An important element of American Legal Realism was the intention to discover how law functioned in actual cases, utilising empirical methods. American Legal Realism is important to this thesis, due to its empirical goals, and the movement is analysed in this chapter.

The majority of ELS research on judicial decision-making is American. American political science research since at least the 1940s has explored the decision-making of judges, leading to the formulation and testing of various 'models' of judicial behaviour. This chapter analyses several models of judicial decision-making which have been tested by ELS researchers. This study does not seek to directly test in a deductive manner any model of judicial decision-making, for the reasons explained in this chapter. Recently, the focus in judicial studies has moved away from behavioural accounts, and towards more nuanced accounts of courts and judicial decision-making. The norms surrounding how a particular court operates, which includes the law orbiting decisions, could have an important role in the decisions of a court. These 'models' of judicial decision-making are

²²³ L Epstein, 'Some Thoughts on the Study of Judicial Behaviour' (2016) 57 Wm & Mary L Rev 2017, 2022.

analysed in this chapter, in order to locate this study within the existing field of judicial studies.

3.1 An explorative positivistic study of judicial decision-making

The study of law and judicial decision-making was at one time known as 'jurimetrics'. The term jurimetrics was coined by Lee Loevinger in the 1940s who described it as 'the scientific investigation of legal problems',²²⁴ such as analysis of the behaviour of witnesses, judges, and legislators.²²⁵ This is indicative of the 'behavioural revolution' of empirical legal studies and political science which was taking place around that time.²²⁶ Loevinger believed that the questions of jurisprudence, such as 'What is law?', had achieved little in two thousand years. Moreover, he felt that the answer to such questions would serve no useful purpose.²²⁷ He contrasted this with the natural sciences, which he thought had achieved a great deal in a relatively short period of time. He wrote that the first task in conducting a scientific inquiry is to 'ask a scientific question ... that can be answered by doing something and observing the result'.²²⁸ He defined jurimetrics as the testing of conclusions about law through utilising the methods of science.²²⁹ This 'testing of conclusions' through scientific methodology is heavily indicative of the positivist deductive paradigm in social science research.

The quantitative ELS process involves moving from a set of questions which are to be answered; to the collection and coding of data which will answer that question; to the analysis of the data to decide whether the initial ideas are reflective of what is happening in the court according to the data.²³⁰ It will be discussed how this is conducted in this study in Chapter 5. The data used to address the question of the Court of Appeal's impartiality are the selected factual,

²²⁴ L Loevinger, 'Jurimetrics: The Next Step Forward' (1949) 33 (5) Minn L Rev 455, 483.

²²⁵ *ibid*, 485-6.

²²⁶ N Maveety (ed) *The Pioneers of Judicial Behavior* (University of Michigan Press 2003) 9.

²²⁷ Loevinger (n 224) at 484.

²²⁸ L Loevinger 'Jurimetrics: The Methodology of Legal Inquiry' (1963) 28 L & Contemp'ry Prob 5, 7.

²²⁹ *ibid*.

²³⁰ L Epstein and AD Martin, *An Introduction to Empirical Legal Research* (Oxford University Press 2014) Chapter 2.

demographic, and legal variables gathered from the Court's judgments. A list of variables is provided in Appendix A, and the variables are evaluated in Chapter 6. The method of collecting the data from the appeals against conviction is a quantitative content analysis, and this is discussed in Chapter 5.

It could be challenged how well positivism as a research paradigm can operate in the social sciences. The core of the challenge is a philosophical 'objection to the idea that one can explain human action in the same way as ... phenomena in the natural world'.²³¹ Interpretivism is one challenge to positivism; this is the claim that social or cultural reality is socially constructed meaning that 'objective' observation is impossible.²³² A further challenge to positivism is from critical scholars such as feminist or critical-race scholars. Those scholars maintain that the positivist or 'objective' version of reality is in fact a male or white-dominated reality and so subject to male and white bias.²³³ Marxist epistemology may make a similar claim against positivism: since the proletariat have particular insights into capitalist society their version of reality may differ from others.²³⁴ Whilst it is accepted that in certain spheres, such as criminal justice, certain concepts such as innocence and guilt are to some extent social constructs, there are some objective facts which can be observed. It is these observable facts which form the variables in this thesis.

Given the method employed in this thesis, a quantitative, objective and value-neutral study of Court of Appeal judgments, a positivistic epistemology is appropriate for the limited section of society which is being studied. There are no claims made about wider reality which need to be reconciled with the challenge of interpretivism; this thesis involves the collection of the information provided in the Court's judgments. It seeks to present a model of the reality of the Court, not

²³¹ See R Banakar and M Travers 'Law, Sociology and Method' in R Banakar and M Travers (eds) *Theory and Method in Socio-Legal Research* (Hart Publishing 2005) 49.

²³² See N Blaikie, *Designing Social Research* (2nd edition, Polity Press 2010) 52.

²³³ See M Hammersley, *The Politics of Social Research* (Sage Publications 1995) 45-6.

²³⁴ *ibid*, 54.

to represent reality itself.²³⁵ The variables have been selected on the basis that they can be discerned with some objectivity and reliability from the Court of Appeal judgments. As will be explained in Chapter 6, some variables could not be collected entirely objectively and so required some interpretation. This is considered in Chapter 6 as being a limitation of the study.

The majority of empirical studies on the decision-making of judges in courts is necessarily quantitative and positivist.²³⁶ A quantitative approach is the most appropriate for the specific questions addressed in this thesis. It is important to note that most of the data collected from the cases is qualitative (such as gender, types of offending, and so on) but is turned into numbers for quantitative analysis. As Hawkins points out, although this study is a quantitative one, it must not lose sight of the social and organisational context in which the decision-making takes place.²³⁷ Indeed, some parts of the 'organisational context' of the Court, form some of the variables in this study. Qualitative social science research does not necessarily turn data into numbers as does quantitative research, but seeks instead to 'document human experience', typically through interviews and fieldwork and the analysis of documents and artefacts.²³⁸ This study draws upon a range of qualitative studies, such as Malleson's²³⁹ and Roberts's²⁴⁰ studies of the approaches of the Court. Darbyshire's study of the working lives of judges provides valuable insights into the culture of decision-making in the Court of Appeal.²⁴¹ Following the analysis of the quantitative data collected on the Court's decisions, there will necessarily be a process of qualitative assessment to contextualise and reflect upon the quantitative elements.

²³⁵ See JA Segal and HJ Spaeth, *The Supreme Court and the Attitudinal Model Revisited*, (Cambridge University Press 2002) 45-6.

²³⁶ K Hawkins 'Order, Rationality and Silence: Some Reflections on Criminal Justice Decision-Making' in L Gelsthorpe and N Padfield (eds) *Exercising Discretion: Decision-Making in the Criminal Justice System and Beyond* (Willan Publishing 2003) 186.

²³⁷ *ibid*, 187.

²³⁸ See J Saldana, *Fundamentals of Qualitative Research*, (Oxford University Press 2011) Chapter 1.

²³⁹ K Malleson, *The Royal Commission on Criminal Justice Review of the Appeal Process*, Research Study No 17 (HMSO 1993).

²⁴⁰ S Roberts, 'The Royal Commission on Criminal Justice and Factual Innocence: Remedying Wrongful Convictions in the Court of Appeal' (2004) (1)(2) JJ 86.

²⁴¹ See P Darbyshire, *Sitting in Judgment: The Working Lives of Judges* (Hart Publishing 2011) Chapter 14.

An important element of positivist research is that it establishes hypotheses which are tested by analysing the data collected from the data source. The analysis employed in this thesis is in the form of hypothesis tests relating to whether the variables selected for analysis are statistically significant predictors of particular outcomes. It is important to note what approach to hypothesis testing is adopted in this study. Within American ELS research, various models of judicial decision-making have been tested, since at least the time of Loewinger. These studies argue that if a certain model is accurate certain effects will be seen in the data. For instance, the 'attitudinal model' argues that if it is true observers will find that judges of different 'ideologies' decide cases differently. Current ELS research, therefore, primarily takes the form of 'hypothetico-deductive confirmatory'²⁴² studies, in which a chosen model is tested by determining whether the postulated effect is seen in the data.

The models of judicial decision-making have rarely been applied to British judges, and so there is no *a priori* reason to suppose that they are any more valid than the legal model. Furthermore, there are reasons, discussed below, to doubt the applicability of American ELS research to the UK. Accordingly, this study is explorative in nature, in order to assess initially whether patterns exist in the data. Some of these patterns are postulated to be 'observable implications' of a possible lack of impartiality, whilst some patterns will be indicative of impartial decision-making. This study may be best considered a descriptive exploratory analysis of any patterns and the extent of those patterns or correlations. This is in keeping with what Tukey said was the attitude of explorative data analysis: 'a willingness to look for what can be seen, whether or not anticipated'.²⁴³ The results are then examined to determine what kinds of conclusions can be drawn from the data regarding the impartiality of the Court of Appeal.

²⁴² This phrase taken from AT Jebb, S Parrgon, and SE Woo, 'Exploratory Data Analysis as a Foundation of Inductive Research' (2017) 27 HR Man Rev 265, 265. Whilst they discuss organisational management research, it is apt for ELS research also.

²⁴³ JW Tukey, 'We Need Both Exploratory and Confirmatory' (1980) 34 Am Stat 23, 24.

3.2 Objectivity

As was explained in Chapter 2, psychological research suggests that a variety of unconscious biases affect many of the ordinary decisions made in everyday life. These biases could also impact decisions made in the process of conducting research. The statistician Regina Nuzzo wrote that ‘failure to understand our own biases has helped to create a crisis of confidence in the reproducibility of published results’.²⁴⁴ By this, she meant that there is a concern that the results of published research may not be able to be reproduced by later researchers because researchers’ biases shape how research is conducted. If biases influence how research is conducted, and people’s biases are different, it may mean that different researchers will find different results. This is problematic if the researcher makes claims, or hopes, to have found objective or general truths, which is indicative of positivist research.

This apparent crisis of confidence has materialised due to the acknowledgement that honest scientists and researchers are subject to confirmation biases which leads them to ‘fixate on collecting evidence to support just one hypothesis; neglect to look for evidence against it; and fail to consider other explanations’.²⁴⁵ This concern surrounding objectivity may lead to doubt whether much of the scientific knowledge about the world which we possess is true. If it is true that researchers search for evidence which supports their conclusions and omit to search for evidence against it, then scientific knowledge is not objective. This concern over the objectivity of knowledge is problematic for those seeking to produce data about the social world within a positivist paradigm.

While it is impossible to exclude the possibility that biases have influenced the research conducted for this thesis, there have been ways to limit bias. Standard statistical procedures have been employed which provide some objective measurement of the strength of variables as predictors. Whilst the value of such measures are only as valid as the data inputted to compute them, the variables

²⁴⁴ R Nuzzo, ‘Fooling Ourselves’ (2015) (526) *Nature* 182, 183.

²⁴⁵ *ibid.*

were specifically selected to reduce the amount of subjectivity needed in the data collection. The decision-making processes involved in collecting the variables, and the specific definitions of particular variables is explained in Chapter 6. The aim behind such comprehensiveness is that it enables replication. Replication is the best way to assess whether conclusions drawn from research are valid and reliable. This thesis has sought to work to a standard to enable replication to take place in later studies.

3.3 American Legal Realism

The majority of empirical studies of law could be considered more or less overtly Legal Realist in perspective.²⁴⁶ Epstein, Landes and Posner take an explicitly Legal Realist position, in that they argued judges have to exercise discretion when determining difficult cases, and may have to fall back on issues of policy.²⁴⁷ The reason for the Realist position in many empirical studies is that the Realists challenged the determinacy of law. The Realists' arguments surrounding the indeterminacy of law was a precursor to the debate between Hart and Dworkin in the 1970s and discussed in Chapter 2, and the Critical Legal Studies movement of the 1980s. Whilst these raised jurisprudential questions, if law has some indeterminacy, there is also an empirical opportunity to discover whether certain factors play a role in decision-making. A key component of the Realist programme was the aim of encouraging empiricalism in law.

As was discussed in Chapter 2, American Legal Realism developed in American Law Schools in the late 1800s and prospered until the 1950s.²⁴⁸ This was in response to the perceived 'formalist' or 'mechanical' age of legal thought said to have existed during the 1800s. It was commonly thought during this period that judicial decision-making was primarily based upon legal logic and analysis of

²⁴⁶ See discussion by PK Chew and RE Kelley 'Myth of the Color-Blind Judge: An Empirical Analysis of Racial Harassment Cases' (2009) *Wash U L Rev* 1117, 1130. See also F Schauer 'Is There A Psychology of Judging?' in D Klein and G Mitchell (eds) *The Psychology of Judicial Decision Making* (Oxford University Press 2010) 110.

²⁴⁷ L Epstein, WM Landes and RA Posner, *The Behaviour of Federal Judges: A Theoretical and Empirical Study of Rational Choice* (Harvard University Press, 2013) 28.

²⁴⁸ See N Duxbury, *Patterns of American Jurisprudence* (Clarendon 1995).

legal texts. The Legal Realist Felix Cohen famously referred to this ‘functional approach’ as being in reliance upon ‘transcendental nonsense’.²⁴⁹ He criticised judges for seeking the answers to legal questions within the legal constructs themselves, rather than by reference to ‘economic, sociological, political, or economic questions’.²⁵⁰

The meaning and philosophy of American Legal Realism is to some extent contested. Karl Llewellyn and Jerome Frank, who could be considered amongst the most prominent Realists, rejected the notion of a ‘school’ of Legal Realism. They said the Realists differed as much amongst themselves as they did from their antithesis, Christopher Columbus Langdell.²⁵¹ According to Twining, part of the difficulty in locating a precise meaning of Legal Realism is that it was so often caricatured or misunderstood itself.²⁵² The Realists’ tendency for flamboyant turns of phrase did not assist understanding or help to avoid this caricature.²⁵³

One element of the Realist critique of judicial decision-making was to demonstrate the indeterminacy and discretion inherent in judicial decision-making. Karl Llewellyn argued that most Realists would agree that ‘*in any case doubtful enough to make litigation respectable the available authoritative premises...are at least two.*’²⁵⁴ When there are at least two authoritative premises, the judge by implication has to choose the most preferable outcome appropriate to the particular facts of the case. The question of what, if it was not the simple application of law, led courts to decide in certain ways was a key question for the Realists. Cohen stated that the element of choice in judicial decision-making meant that ‘the study of social factors that determine the course

²⁴⁹ FS Cohen, ‘Transcendental Nonsense and the Functional Approach’ (1935) Colum L Rev 809.

²⁵⁰ *ibid*, 810.

²⁵¹ KN Llewellyn ‘Some Realism about Realism: Responding to Dean Roscoe Pound’ (1931) (44) Harvard L Rev 1222, 1233-4. Although the article is only in Llewellyn’s name, he states that it was written by him and Frank.

²⁵² W Twining, ‘Talk About Realism’ (1985) (60)(30) NYU L Rev 329.

²⁵³ See Introduction to KN Llewellyn, *The Theory of Rules* (edited and compiled by F Schauer) (University of Chicago Press 2011).

²⁵⁴ Llewellyn (n 251 above) 1239.

of judicial decision' should be a key skill of lawyers in order to be able to predict how judges would decide a particular case.²⁵⁵

Jerome Frank argued that what constituted the law of a particular case was the legal rules and the judge's perception of the facts.²⁵⁶ As perceptions of facts can differ from person-to-person the application of the law to the facts was essentially discretionary. Frank said discretion and law 'cannot be separated ... They are so thoroughly intermingled that it is impossible to divide them'.²⁵⁷ Frank's particular version of Realism²⁵⁸ placed the situation of a judge as a human being at the core. Contested cases were not decided syllogistically, but instead:

'The judge's knowledge of the rules combines with his reactions to the conflicting testimony, with his sense of fairness, with his background and economic and social views, and with that complicated compound loosely named his "personality", to form an incalculable mixture out of which comes the court order we call his decision'.²⁵⁹

In this sense, 'the Realists saw legal reasoning as human reason, and traditional claims for the distinctiveness of legal reasoning were largely pretence'.²⁶⁰ The Realists' arguments regarding the decision-making of judges form the foundations of the behavioural and attitudinal models which are discussed below. The 'discovery', however, that the law often points in both ways, and that judges act as human beings, could hardly be considered a revolutionary finding of the Realists. As Tamanaha explains, this 'realism' about judging, in the observation

²⁵⁵ LK Cohen (Ed) *The Legal Conscience: Selected Papers of Felix S Cohen* (Yale University Press 1960) 84

²⁵⁶ See J Frank, 'Are Judges Human? Part 1: The Effect on Legal Thinking of the Assumption that Judges Behave Like Human Beings' (1932) 80 U Pa L Rev 17, 22.

²⁵⁷ *ibid.*

²⁵⁸ Which may be considered a more extreme kind, see B Leiter, *Naturalizing Jurisprudence: Essays on American Legal Realism and Naturalism in Legal Philosophy*, (Oxford University Press 2007) 17.

²⁵⁹ Frank (n 256) at 47.

²⁶⁰ F Schauer 'Is There A Psychology of Judging?' in D Klein, G Mitchell (eds) *The Psychology of Judicial Decision Making* (Oxford University Press 2010) 110.

that they have to make choices, was commonplace before the Realists.²⁶¹ Epstein and her colleagues claimed that the first full articulation of a 'realist' conception of judging was made by Jeremy Bentham in the 1780s.²⁶² Posner located Realism in the writing of Plato.²⁶³ Tamanaha argued that Legal Realism should not be understood as a movement beginning at a distinct point in time, around the time of Holme's article discussed in Chapter 1, but as a generally shared perspective to law since at least the 1800s.²⁶⁴ Nevertheless, the collective of lawyers and legal scholars known as the American Legal Realists are important because they set in place the implantation of the social sciences into American law schools.²⁶⁵

According to Leiter, a core claim of most scholars searching for a realistic study of law was that they thought 'judges respond primarily to the stimulus of facts', in that 'judges reach decisions based on what they think would be fair on the facts of the case'.²⁶⁶ Leiter argued that the Realists believed law was 'locally' indeterminate, at least at the stage of appellate review, where law did not determine one right answer.²⁶⁷ As seen above, some Realists, such as Frank, suggested that the indeterminacy within law was greater than this.²⁶⁸ This more radical, or 'Frankified'²⁶⁹ version of Realism may have ultimately led to the decline of Legal Realism as a legal philosophy. This is because upon a 'Frankified' version of Legal Realism it becomes impossible to predict the decision-making of judges owing to the discretionary nature of law on that account. Indeed, Frank stated that due to the human element of judicial decision-making, prediction was impossible.²⁷⁰ The difficulty with this version of Realism is that many laws do

²⁶¹ BZ Tamanaha 'Legal Realism in Context' in E Mertz, S Macaulay and TW Mitchell (eds), *The New Legal Realism: Translating Law-and-Society for Today's Legal Practice* (Cambridge University Press 2016) 156-61.

²⁶² L Epstein, WM Landes and RA Posner, *The Behaviour of Federal Judges: A Theoretical and Empirical Study of Rational Choice* (Harvard University Press, 2013) 27.

²⁶³ RA Posner, 'Realism About Judges' (2011) *Nw U L Rev* 577, 577.

²⁶⁴ Tamanaha (n 261 above) at 164.

²⁶⁵ *ibid*, 165.

²⁶⁶ B Leiter, *Naturalizing Jurisprudence: Essays on American Legal Realism and Naturalism in Legal Philosophy*, (Oxford University Press 2007) 22.

²⁶⁷ *ibid*, 41.

²⁶⁸ J Frank, *Law and the Modern Mind* (Transaction Publishers 2009).

²⁶⁹ Leiter (n 266) at 17

²⁷⁰ See Frank, (n 268 above) at e.g. 23.

operate in a predictable way, and lawyers do predict the decision-making of judges, by, for example, deciding not to continue with litigation.

Llewellyn's more mainstream version of Realism focussed upon appellate review.²⁷¹ Llewellyn argued that judges' decisions are predictable because there were various '[stabilising] factors'.²⁷² Llewellyn stressed that appeal judges are human beings who are 'law-conditioned'²⁷³ and guided by 'legal doctrine'.²⁷⁴ Llewellyn argued that it was especially at the stage of appellate review the 'interpretation of language or the sizing up of the facts, or the choice open as among available divergent premises ... will allow a fair technical case to be made either way'.²⁷⁵ It may be observed that a mainstream Realist conception of judicial decision-making is similar to Lord Bingham's discussed in Chapter 2, and, as discussed below, may in fact be largely compatible with Hart's.²⁷⁶

The American Legal Realists held empirical goals in order to observe what happened when the law did not determine the outcomes of litigation. Llewellyn discussed the first attempts to '[capitalise] the wealth of our reported cases to make large-scale quantitative studies of facts and outcome'. At the time Llewellyn wrote those words, however, he knew of 'no published results'.²⁷⁷ The Realists did not generally produce a great deal of empirical work themselves. Underhill Moore was one of the few in the centre of the Realist movement to conduct empirical studies of the law in action. His study of the effect of parking regulations on the behaviour of drivers in New Haven is his most famous work.²⁷⁸ Moore's work was criticised and ridiculed as being trivial at the time,²⁷⁹ but has been more

²⁷¹ See KN Llewellyn, *The Common Law Tradition: Deciding Appeals* (Little, Brown & Co 1960).

²⁷² *ibid.*, 19.

²⁷³ *ibid.*

²⁷⁴ *ibid.*, 20.

²⁷⁵ *ibid.*

²⁷⁶ See TH Hunter, 'HLA Hart's Concept of Law in the Context of American Legal Realism' (1972) 35 MLR 606.

²⁷⁷ KN Llewellyn 'Some Realism about Realism: Responding to Dean Roscoe Pound' (1931) (44) Harvard L Rev 1222, 1244.

²⁷⁸ U Moore and CC Callahan 'Law and Learning Theory: A Study in Legal Control' (1943) (53)(1) Yale LJ 1.

²⁷⁹ See FSC Northrop 'Underhill Moore's Legal Science: Its Nature and Significance' (1950)(59) Yale LJ 196.

charitably interpreted in recent years as an exemplar of law-in-action empirical scholarship.²⁸⁰ As a result of failing to conduct empirical research themselves, the Realists challenged the determinacy of law, but did not explain what filled the gap. This meant, as Cross says, Realism took on a nihilistic look.²⁸¹

Perhaps as a result of this, American Legal Realism flourished during the inter-War years, but had 'ran itself into the sand' by the 1950s.²⁸² One reason for this may have been that some Realists' apparent scepticism towards legal rules led them to moral relativism or nihilism, which was not an attractive prospect after the Second World War.²⁸³ They could not provide an answer to what fills the gaps in law. Part of the reason for this may have been a lack of funding during the 1930s Great Depression or, according to Schlegel, the 'peculiarities of the personalities' of the leaders of the movement, or the protagonists simply 'losing their nerve'.²⁸⁴ White wrote that the Realists' process of 'debunking' followed by 'objective' observation was a contradiction in terms.²⁸⁵ 'Debunking' stemmed from the seminal work of Hohfeld who sought to show that certain legal concepts are often indiscriminately used to convey what in fact is its 'opposite' or its 'correlative'.²⁸⁶ But at the same time as 'debunking' functionalist claims of legal objectivity, the Realists assumed that their (limited) empirical observations would be value-free.²⁸⁷

American Legal Realism received its final blow in 1961 with Chapter 7 of Hart's *The Concept of Law*.²⁸⁸ Hart attributed to the Realists the claim that 'talk of rules

²⁸⁰ DE Ho and DB Rubin, 'Credible Causal Inference for Empirical Legal Studies' (2011) (7) *Ann Rev of Law and Soc Sci* 17.

²⁸¹ FB Cross, 'The New Legal Realism and Statutory Interpretation' (2013) (1)(1) *T & P of Legislation* 129, 131.

²⁸² JH Schlegel, 'American Legal Realism and Empirical Social Science: From the Yale Experience' (1979) (28) *Buff L Rev* 459.

²⁸³ GE White, 'From Realism to Critical Legal Studies: A Truncated Intellectual History' (1987) (40) *SW LJ* 819, 824.

²⁸⁴ See Schlegel (n 282 above) 460.

²⁸⁵ See White (n 283 above) 821-3.

²⁸⁶ WN Hohfeld 'Some Fundamental Legal Conceptions As Applied in Judicial Reasoning' (1914) (23) *Yale LJ* 16, 30.

²⁸⁷ See White (n 283) at 823.

²⁸⁸ HLA Hart, *The Concept of Law* (3rd edition, Oxford University Press 2012).

is a myth', and that 'law consists simply of the decisions of courts and the predictions of them'.²⁸⁹ It will be recalled from Chapter 1 that Holmes referred to law as being the prediction of what judges will do.²⁹⁰ This could not be true, Hart argued, because judges are guided by the law when they make decisions.²⁹¹ Hart accused the Realists of sometimes being a 'disappointed absolutist'²⁹² since, having discovered the 'open-texture' at the fringes of law they exaggerated its effect so as to make this indeterminacy appear part of the concept of law.²⁹³ This also could not be true because many rules of law, such as speed limits, or rules relating to the drawing of wills, are routinely applied. Hart's critique is said by Leiter to have rendered Legal Realism a 'philosophical joke'.²⁹⁴ As Green says, it is a cliché to say that 'we are all realists now,' but it is also mistaken; legal philosophers are rarely legal realists.²⁹⁵

Since then, Hart's alleged misreading of the Realists has been well documented.²⁹⁶ Leiter claims that Hart was asking a jurisprudential question about the concept of law, whilst the Realists attempted to provide a description and analysis of adjudication.²⁹⁷ According to Leiter, the Legal Realist critique was lawyerly rather than philosophical, avoiding the inevitable descent into nihilism. Many of the Legal Realists were judges themselves (for instance Holmes; Frank; and Hutcheson), or senior members of the legal establishment (such as Llewellyn, who drafted the Uniform Commercial Code), rather than philosophers, which does not coalesce with the extreme scepticism about judging which Hart

²⁸⁹ *ibid*, 136.

²⁹⁰ OW Holmes, 'Path of the Law' (1897) 1 Boston L School Mag 1

²⁹¹ See Hart (n 288 above) 136-9.

²⁹² *ibid*, 139.

²⁹³ *ibid*, 126.

²⁹⁴ B Leiter, *Naturalizing Jurisprudence: Essays on American Legal Realism and Naturalism in Legal Philosophy* (Oxford University Press 2007) 20.

²⁹⁵ MS Green 'Legal Realism as a Theory of Law' (2005) 46(6) W & M L Rev 1915.

²⁹⁶ See Leiter (n 294); H Dagan, *Reconstructing American Legal Realism and Rethinking Private Law* (Oxford University Press 2013); F Schauer 'Legal Realism Untamed' (2013) Tex LR 749; W Twining, *Karl Llewellyn and the Realist Movement* (2nd ed, Cambridge University Press 2012) 406-7.

²⁹⁷ See Leiter (n 294), Chapter 1.

attributed to the Realists.²⁹⁸ Llewellyn himself stated that Realism 'is not a philosophy, but a technology', or a method, of 'see it fresh'.²⁹⁹

The importance of American Legal Realism on the American legal academy, and for this thesis in particular, is that it gave rise to the principle of conducting empirical research on courts.³⁰⁰ As Cross argued, Legal Realism may best be understood as a mood within the legal academy which sought a 'realism' about judging and legal processes.³⁰¹ The purpose of this thesis is to determine whether the variables included in the analysis are predictors of the outcome of appeals against conviction. The research conducted in this thesis is an example of ELS research, which claims a direct link to the original Legal Realists.³⁰² Thus, Legal Realism provides a jurisprudential and historical lineage to the empirical analysis of judges. Within ELS research, a variety of 'models' of judicial decision-making have been developed. This thesis works within this literature, and variables have been drawn from it. These models are discussed below.

3.4 New Legal Realism

While the American Legal Realists held empirical goals but rarely acted upon them, modern legal scholarship contains an abundance of empirical studies on the behaviour of judges. Twining argued that the debate surrounding the meaning of American Legal Realism should now be considered legal history, and 'New Legal Realism' should free itself from such debates.³⁰³ New Legal Realism differs from ELS in that ELS tends to emphasise quantitative research on the higher judiciary, whilst New Legal Realism proposes a 'big tent' approach, incorporating a wider range of approaches, including qualitative and other social-

²⁹⁸ See BZ Tamanaha 'Legal Realism in Context' in E Mertz, S Macaulay, TW Mitchell (eds) *The New Legal Realism: Translating Law-and-Society for Today's Legal Practice* (Cambridge University Press 2016) 156.

²⁹⁹ KN Llewellyn, *The Common Law Tradition: Deciding Appeals* (Little, Brown & Co 1960) 510.

³⁰⁰ See CG Geyh, 'Can the Rule of Law Survive Judicial Politics?' (2012) 97 *Corn L Rev* 191, 201.

³⁰¹ Cross (n 281) 130.

³⁰² See TJ Miles and CR Sunstein, 'The New Legal Realism' (2008) 75 *U Chic L Rev* 831.

³⁰³ W Twining, 'Legal R/realism and Jurisprudence: Ten Theses' in E Mertz, S Macaulay, and TW Mitchell (Eds) *The New Legal Realism: Translating Law and Society for Today's Legal Practice* (Vol 1) (Cambridge University Press 2016).

science research methods, such as ethnography.³⁰⁴ New Legal Realism appears to prefer a 'bottom-up' approach, with a focus on 'how law actually works in people's lives'.³⁰⁵ Similar to the original Legal Realists, ELS has a 'top-down' approach with its emphasis on appellate decisions. Mertz argued that while New Legal Realism's 'big tent' approach 'would welcome the insights of researchers into judicial behaviour' they would also insist upon such researchers 'taking other forms of disciplinary knowledge into account'.³⁰⁶ Miles and Sunstein argued that ELS might best be considered as a 'subpart of a broader conception of the New Legal Realism'.³⁰⁷

There have now been many empirical studies of the outcomes of cases with a view to explaining judicial behaviour, and this appears to be growing.³⁰⁸ The variables utilised in these earlier ELS studies have driven the selection of variables in this thesis. These studies have been largely confined to the United States, reflecting the great impact on Legal Realism on American scholarship. The work of the Legal Realists and the New Realists is not complete, as Miles and Sunstein say:

'We continue to know only a small amount about what might be learned with respect to the effects of key aspects of judicial background on judicial voting... How do sex and race affect [behaviour] in multiple areas of the law? Are female appointees more likely to be pro-choice? In these domains, we glimpse only the tip of the iceberg.'³⁰⁹

This statement was made from the perspective of American ELS scholars who carry out empirical work on judges. In the United Kingdom, the range of empirical study of the decision-making of judges is far less extensive. The *Nuffield Enquiry*

³⁰⁴ See Mertz, 'Introduction' in Mertz, Macaulay and Mitchell, *ibid*, at 4.

³⁰⁵ *ibid*, 7.

³⁰⁶ *ibid*, 5-6.

³⁰⁷ Miles and Sunstein (n 302) at footnote 16 in text.

³⁰⁸ M Heise 'An Empirical Analysis of Empirical Legal Scholarship Production, 1990-2009' (2011) *U III L Rev* 1739.

³⁰⁹ Miles and Sunstein, (n 302) 842.

on *Empirical Legal Research* expressed concern at the limited number of empirical studies on law carried out in the United Kingdom.³¹⁰ As Adler and Simon suggested, however, when the Enquiry referred to there being a lack of interest in ‘empirical research’, they appear to have meant more specifically ‘quantitative empirical research’, in particular quantitative empirical research of judges.³¹¹ The Enquiry may have overlooked the many socio-legal empirical studies on the United Kingdom legal system (such as the police and elements of policing, and the jury system).³¹² Many of these, such as McBarnet’s research from the early 1980s on the operation of safeguards to protect people suspected of crime,³¹³ McConville, Sanders, and Leng’s study into the operation in practice of the *Police and Criminal Evidence Act 1984*, and McConville’s study of criminal defence lawyers,³¹⁴ could well be considered classic examples of New Legal Realism, given their examinations of the operation of law in everyday life and their critical evaluations of the balance of power in society.

The *Nuffield Enquiry* suggested there was a lack of ‘critical mass’ of researchers conducting empirical research in the United Kingdom,³¹⁵ and it is submitted that this is particularly true for large-*n* quantitative studies of judges. One likely reason that there is not a culture of quantitative studies of judges in the UK is that the influence of ‘old’ Legal Realism was far more limited in the UK than it was in the US. There was accordingly no turn-of-the-20th-century ‘growth spurt’ in empirical research as there was in the US.³¹⁶ In the UK there have been relatively few

³¹⁰ H Genn, M Parlington and S Wheeler, *The Nuffield Enquiry on Empirical Legal Research* (Nuffield Foundation 2006).

³¹¹ M Adler and J Simon, ‘Stepwise Progression: The Past, Present, and Possible Future of Empirical Research on Law in the United States and United Kingdom’ (2014) 41 *J Law & Soc* 173, 191.

³¹² See for instance, amongst many others, M McConville, A Sanders and R Leng, *The Case for the Prosecution* (Routledge 1991); JS Baxter, SA Bain and JH McAusland ‘“Are You Thick?”: A Case Study of Confrontational Aspects Of A Police Interview With A Suspect’ (2007) *PJ* 99; L Skinnis ‘I’m a detainee; Get Me Out of Here’ (2009) 49 *Brit J Criminol* 399; C Thomas, *Are Juries Fair?* (MOJ 2010); V Kemp, P Pleasence and NJ Balmer ‘Children, Young People and Requests for Police Station Legal Advice: 25 Years On From PACE’ (2011) *Youth Justice* 11.

³¹³ DJ McBarnet, *Conviction: The Law, The State and the Construction of Justice* (Palgrave Macmillan 1981).

³¹⁴ M McConville, *Standing Accused: The Organisation and Practices of Criminal Defence Lawyers in Britain* (Clarendon 1994).

³¹⁵ *Nuffield Enquiry* (n 310).

³¹⁶ See Adler and Simon (n 311).

quantitative empirical studies of the decision-making of judges. One such study is Robertson's study of the House of Lords.³¹⁷ He presented a statistical model which was able to predict the outcomes of cases, at times with over 90% accuracy, based only upon which judges were hearing the case.³¹⁸ Robertson studied 'nearly all' reported cases heard before the House of Lords between 1986 and 1995.³¹⁹ These were coded into SPSS with the judges themselves as the independent variables to determine the impact of particular judges on the likelihood of an appeal being successful. He found, for instance, that in tax cases the taxpayer was substantially more likely to lose if Lord Templeman was on the bench.³²⁰

Large-scale empirical scholarship of the decision-making of judges, drawing upon Legal Realism, accelerated during the 1960s and considered 'behavioural' explanations for judicial decisions. This has developed to the present day with numerous other models of judicial decision-making, such as the attitudinal model, strategic model, and institutional models. These models are not directly tested in this thesis but the variables analysed are derived from these models. The next section analyses the study of judicial behaviour, before stating the place of this study within this literature.

3.5 Behavioural models of judicial decision-making

In the 1960s, Schubert emphasised the human aspect to judicial decision-making: judges are human beings who happen to be cast in important adjudicatory roles.³²¹ Thus, 'judicial behaviouralism'³²² research considers the role of judges' social background or personal attributes in judicial decision-

³¹⁷ D Robertson, *Judicial Discretion in the House of Lords* (Clarendon Press 1998).

³¹⁸ *ibid*, 50-55.

³¹⁹ *ibid*, 40-41.

³²⁰ *ibid*, 37.

³²¹ G Schubert 'Behavioral Jurisprudence' (1967) 2 *Law & Soc'y Rev* 407, 409-10.

³²² In this thesis, the British spelling 'behaviourism' and 'behaviouralism' is adopted. Note, however, that many behaviouralist studies of judges are American, where the word is spelled 'behavioralism'.

making.³²³ Judicial behaviouralism is the theory that social background and personal attributes 'shape personal and policy values that directly influence judicial decisions'.³²⁴ Many of the studies to be discussed in Chapter 5 have influences of judicial behaviouralism. One study in particular has influenced the methods of this study: Sisk, Heise, and Morriss's 1998 '*Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning*'.³²⁵ This study is analysed further in Chapter 5.

The personal characteristics of judges have usually been found to be only mildly correlated with judicial decisions.³²⁶ One of the criticisms of behavioural research is that it may be too simplistic a method to seek to uncover important facts about judicial decision-making. As Clayton and Gillman assert: 'there are important questions left unanswered by narrowly focussing on how particular justices vote in discrete cases'.³²⁷ It is for this reason that they prefer a method which looks beyond personal characteristics, such as the institutional model, discussed below.

In this thesis, a range of variables which could be considered behavioural are considered. A list of the variable is provided in Appendix A. Some of these variables are personal characteristics pertaining to judges, appellants, and complainants / deceased. A full discussion of all variables used is provided in Chapter 6. Examples of the personal characteristics analysed in this study are: who the individual judges hearing the appeal were; the gender of judges, lawyers, and complainants / the deceased;³²⁸ whether the appellant was of previous bad or good character, and whether appellant was under 18 at the time of the allegations. These are not all the variables used in the study, and, indeed, not all

³²³ See M Heise 'The Past, Present, and Future of Empirical Legal Scholarship: Judicial Decision Making and the New Empiricalism' (2002) U Ill L Rev 819, 833.

³²⁴ GC Sisk, M Heise, and AP Morriss 'Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning' (1998) (73) NYUL Rev 1377, 1385. (Emphasis added).

³²⁵ *ibid.*

³²⁶ M Heise (n 323 above) at 834.

³²⁷ C Clayton and H Gillman in Gillman and Clayton (eds) *The Supreme Court in American Politics: New Institutional Interpretations* (University Press of Kansas 1999) 1.

³²⁸ The appellants in this study were almost always male.

variables which could have been used have been used. As discussed in Chapters 5 and 6, the omission of certain variables is a limitation of this study. However, it is important to emphasise, as Whittington suggested, that it is not the intention of this thesis to show that behaviouralism, or any other model, is 'right'.³²⁹ Rather, it is an initial explorative study of whether a range of variables are predictors of the outcome of appeals against conviction, which is argued to be an adequate if incomplete measurement of impartiality.

The behavioural model of judicial decision-making has been extended upon by the attitudinal model. The attitudinal model was developed by Pritchett, who observed changing degrees of dissent in the US Supreme Court, and saw that divisions arose between those said to be on the left and those on the right of the Court.³³⁰ Pritchett's work was extended 'in terms of depth, sophistication and precision'³³¹ by Schubert in the late 1950s and 1960s.³³² The attitudinal model has since been developed by Professors Segal and Spaeth following analysis of the US Supreme Court Judicial Database (USSCJD) which they developed.³³³ The attitudinal model maintains that judges' attitudes or ideology 'is a complete and adequate model of the Supreme Court's decisions on its merits' and that the attitudes of judges 'are all that systematically explain the votes of the justices'.³³⁴ As was discussed in Chapter 2, they envisaged the attitudinal model as being a direct challenge to the legal model.

Whilst the behavioural model of judicial behaviour has not gained high empirical support, the attitudinal model is dominant in the literature.³³⁵ Studies have

³²⁹ KE Whittington, 'Once More Unto the Breach: PostBehavioralist Approaches to Judicial Politics' (2000) (25) L & Soc Inquiry 601, 628-9.

³³⁰ CH Pritchett, 'Divided Supreme Court, 1944-1945' (1945) 44 Mich L Rev 427.

³³¹ See NL Maveety (Ed), *The Pioneers of Judicial Behavior* (University of Michigan Press 2003) 14

³³² See GA Schubert, *Quantitative Analysis of Judicial Behavior* (Free Press 1959); GA Schubert, *The Judicial Mind: Attitudes and Ideologies of Supreme Court Justices 1946 – 1963* (Northwestern University Press 1965).

³³³ See <http://supremecourtdatabase.org/> and JA Segal and AD Cover 'Ideological Values and the Votes of US Supreme Court Justices' (1989) (83) Am Pol Sci Review 557.

³³⁴ See JA Segal and HJ Spaeth, 'Response to Critics', (1994)(4) Law and Courts 10, 11.

³³⁵ M Heise (n 323 above) at 833.

frequently found a relationship between ideology and judicial decisions.³³⁶ The usual proxy for ideology is the party of the President who appointed the judge.³³⁷ For instance, Cross and Tiller³³⁸ sought to determine whether judges were applying the US Supreme Court 'deference doctrine' espoused in *Chevron*³³⁹ in the way intended by the Supreme Court, or whether judges were 'manipulating the deference doctrine to achieve politically desirable outcomes'.³⁴⁰ They tested this theory by studying and coding all DC Circuit Court of Appeals cases which cited *Chevron* between 1991 and 1995. Two of the variables they coded were whether the outcome was liberal or conservative, and the partisan makeup of the Court panel (Democrat (liberal) or Republican (conservative)).³⁴¹ This second variable was determined by the party of the President who appointed the judge. They found that a panel consisting of a majority of Republicans rendered more conservative decisions and a panel containing Democrats rendered more liberal decisions.³⁴²

Similarly, Revesz sought to test whether the policy preferences of judges had an impact on how they voted in environmental cases.³⁴³ The policy preference of the judge was also captured by recording the views generally held by the party of the President which appointed the judge.³⁴⁴ His three central conclusions were that ideology influences decision-making; ideological voting is more prevalent in cases less likely to be appealed to the US Supreme Court; and the votes cast differed depending upon the ideology of other judges.³⁴⁵ Sunstein, Schkade and Ellman employed the political affiliation of the appointing President as a proxy for judges' attitudes.³⁴⁶ They found an association between the party which

³³⁶ *ibid*, 837.

³³⁷ *ibid*.

³³⁸ FB Cross and EH Tiller, 'Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals' (1998) 107 Yale LJ 2155.

³³⁹ *Chevron USA Inc v Natural Resources Defense Council Inc* (1984) 467 US 837.

³⁴⁰ Cross and Tiller (note 338) 2162.

³⁴¹ *ibid*, 2168.

³⁴² *ibid*, 2169.

³⁴³ RL Revesz 'Environmental Regulation, Ideology, and the DC Circuit' (1997)(83) Virginia Law Review 1717.

³⁴⁴ *ibid*, 1718.

³⁴⁵ *ibid*, 1719.

³⁴⁶ CR Sunstein, D Schkade and LM Ellman 'Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation' (2004) (90) Virginia L Rev 301, 303.

appointed the judge and decision-making in many areas of law including sex discrimination, sexual harassment and disability discrimination. Interestingly, the hypothesis of ideological voting was rebutted in criminal appeals.³⁴⁷ It was suggested by them that this may be because criminal laws are clearer than other laws, or judges of different 'ideology' have fewer differences in criminal cases.³⁴⁸

The attitudinal model has been criticised by American judges who claim that the model does not reflect their own experience of judging,³⁴⁹ and from scholars on the grounds of its methodology.³⁵⁰ As Fischman and Law point out, it is difficult to know whether the party of the appointing President is a valid measurement of ideology.³⁵¹ It may be that the assumption that Democrat judges have liberal ideologies, and Republican appointed judges have conservative ideologies does not always hold.³⁵² Segal and Cover sought to enhance the measurement of ideology by searching for outside evidence of judicial preferences.³⁵³ They content analysed leading newspaper editorials for expert opinion on the ideological preferences of US Supreme Court justices. This gave each judge a Segal / Cover score of a conservative, moderate, or liberal ideology. The Segal / Cover method has since been utilised frequently by political scientists, but, as Epstein and Mershon noted, later uses of the scores may have stretched the scores beyond their originally intended purposes.³⁵⁴ More recently, researchers have sought to modify the Segal / Cover scores by including quantitative assessment of the legislature's ideology at the time a judge was appointed.³⁵⁵

³⁴⁷ *ibid*, 334.

³⁴⁸ *ibid*, 334-5.

³⁴⁹ See PM Wald, 'A Response to Tiller and Cross' (1999) 99 *Columbia L Rev* 235; HT. Edwards, MA. Livermore 'The Pitfalls of Empirical Studies that Attempt to Understand the Factors Affecting Appellant Decisionmaking' (2009) (58)(8) *Duke LJ* 1895.

³⁵⁰ See L Epstein and G King 'The Rules of Inference' (2002) 69 *U of Chic L Rev* 1, 88-9.

³⁵¹ JB Fischman, and DS Law 'What is Judicial Ideology, and How Should We Measure It?' (2009) (29) *Wash U J of Law & Pol'y* 133

³⁵² See L Epstein and G King, 'The Rules of Inference' (2002) 69 *U Chic I Rev* 1

³⁵³ JA Segal and DC Cover, 'Ideological Values and the Votes of Supreme Court Justices' (1989) 83 *Am Pol Sci Rev* 557.

³⁵⁴ L Epstein and C Mershon, 'Measuring Political Preferences' (1996) 40 (1) *Am J Pol Sci* 261.

³⁵⁵ See CM Cameron and J-K Park, 'How Will They Vote? Predicting the Future Behaviour of Supreme Court Nominees, 1937-2006' (2009) 6 (3) *JELS* 485.

The methods of behavioural / attitudinal research have been heavily criticised by some.³⁵⁶ Despite this, the empirical evidence that judges appointed by different Presidents (that is, without the imputation that this demonstrates ‘ideology’) decide differently is strong.³⁵⁷ The attitudinal model does not appear to be an appropriate model to test deductively in this study, because, even if it is appropriate for studies in the United States, it does not appear applicable in Britain. Although the US Constitution has a strong separation of powers ethos, and values highly judicial independence, this appears to be achieved via a ‘checks and balances’ system rather than separation, at least in relation to appointment of judges. For instance, justices of the US Supreme Court are nominated by the serving President and the nominee is confirmed by a simple majority vote in the Senate. Thus, the President cannot override Congress in his selection of judges. The Chief Justice is appointed in the same way.³⁵⁸ Thus, while the judges are appointed this is checked and balanced by Congress.

In the UK, there is no political vote on the appointment of judges, meaning that it is difficult to locate a suitable proxy for judicial ideology. Since 2006, the Judicial Appointments Commission (JAC) has been independently tasked with appointing judges, including the Lord Chief Justice. The independence of the judiciary from the other branches of government is confirmed by the *Constitutional Reform Act 2005*,³⁵⁹ which created the JAC,³⁶⁰ and effected the symbolic renaming of the Appellate Committee of the House of Lords to the Supreme Court.³⁶¹ Hanretty examined whether the outcomes of cases in the House of Lords could be modelled according to the party in power at the time individual judges were appointed.³⁶² He found that such a model did not improve upon the null model as a predictor of outcomes. The political inclinations of judges are rarely disclosed

³⁵⁶ L Epstein and G King, ‘The Rules of Inference’ (2002) 69 U Chic L Rev 1; HT Edwards and MA Livermore, ‘Pitfalls of Empirical Studies that Attempt to Understand the Factors Affecting Appellate Decision Making’ (2009) (58) (8) Duke LJ 1895.

³⁵⁷ See RA Posner, ‘The Role of the Judge in the Twenty-First Century’ (2006) BUL Rev 1049.

³⁵⁸ See <http://www.supremecourt.gov/faq.aspx#faqj1> <accessed 20 August 2015>.

³⁵⁹ Section 3.

³⁶⁰ *ibid*, s 61.

³⁶¹ *ibid*, s 23.

³⁶² C Hanretty ‘The Decisions and Ideal Points of British Law Lords’ (2013) 43 (3) Brit J of Pol Sci 703

in public, meaning it is difficult to find any other way to measure ideology. As Gee and his colleagues say, the appointment of judges in the UK, which was once characterised by ‘stability, secrecy and informality’,³⁶³ is now characterised by a long and highly formalised selection process administered by the JAC. Although it was intended to be a recommending body, it now operates as an appointing body,³⁶⁴ with Lord Chancellors now seriously marginalised even for the more senior judicial positions.³⁶⁵

Arvind and Stirton offered an alternative way of measuring judicial attitude in the UK Supreme Court.³⁶⁶ Rather than analysing judges on a left / right continuum as did Hanretty, they used a green-light / red-light continuum. This, they argued, measured ‘whether they see the proper role of the courts as taking respectively a restrictive or permissive’ attitude towards administrative discretion.³⁶⁷ This was not a purely attitudinal measure, however, but was also a doctrinal / legal measure. This is because they argued that their model captured whether the judges had certain attitudes towards what the law required.³⁶⁸ They found that some judges were more ‘pro-State’ (i.e. green-light) than others, and, at least sometimes, the ‘bench composition can still matter to how cases are decided’.³⁶⁹ Arvind and Stirton have valuably shown that there are alternative measures of judicial attitudes and perspectives, and they can be used to study British judges. Owing to the nature of Court of Appeal judgments, in particular, that there is only ever one opinion in criminal appeals, there is no ideological ranking of judges in this study. However, individual judges are used as variables, which will give an insight into whether individual judges are associated with particular outcomes.

³⁶³ G Gee, R Hazell, K Malleson and P O’Brien, *The Politics of Judicial Independence in the UK’s Changing Constitution*, (Cambridge University Press 2015) 160.

³⁶⁴ *ibid*, 163.

³⁶⁵ *ibid*, 186.

³⁶⁶ TT Arvind and L Stirton, ‘Legal Ideology, Legal Doctrine and the UK’s Top Judges’ (2016) PL 418.

³⁶⁷ *ibid*, 425-6.

³⁶⁸ *ibid*.

³⁶⁹ *ibid*, 429.

It is becoming increasingly acknowledged that behavioural or attitudinal models of judicial decision-making omit some important features of adjudication. In more recent years, judicial scholars and political scientists have sought broader analysis. Of particular importance to this study is work which considers the institutional position of courts, to which attention now turns.

3.6 Post-attitudinal models of judicial decision-making

Gillman and Clayton argued that attention should be diverted away from the policy preferences of particular judges, and towards the 'distinctive characteristics of the court as an institution, its relationship to other institutions, and how both of these might shape judicial values'.³⁷⁰ Bloom argued that legal scholarship entered a 'post-attitudinal' phase with the development of the institutional model, which requires looking 'beyond judicial attitudes'.³⁷¹ As Cross argued, one of the shortcomings of behavioural / attitudinal studies of judicial decision-making is the lack of any variable designed to capture the effects of the law on judicial decisions.³⁷² A consideration of the law is one important element of post-attitudinal research on judicial decision-making. A major criticism of American ELS which utilises the US Supreme Court Database³⁷³ is that it fails to take account of the law itself.³⁷⁴ In this thesis, variables which consider the broader institutional situation and relationships between the Court and other institutions are considered.

As will be explained further in Chapter 4, the decision-making of the Court, and the 'unsafety test' is a product of its history and its place within the criminal justice system. Thus, certain norms may have an impact on how the law is interpreted, and how the Court decides cases. The norms which operate in the decision-making of the Court of Appeal have been previously studied by researchers.

³⁷⁰ CW Clayton and H Gillman (eds) *Supreme Court Decision-Making: New Institutional Approaches* (University of Chicago Press 1999) 3.

³⁷¹ A Bloom 'The "Post-Attitudinal Moment": Judicial Policymaking Through the Lens of New Institutionalism' (2001) (35) *Law & Soc'y Rev* 219, 220.

³⁷² FB Cross, *Decision Making in the US Courts of Appeals* (Stanford University Press 2007) 39.

³⁷³ Housed at Washington University Law School, available at <http://scdb.wustl.edu/>

³⁷⁴ C Shapiro, 'Coding Complexity: Bringing Law to the Empirical Analysis of the Supreme Court' (2009) *Hastings LJ* 477.

Some observers of the Court of Appeal are of the view that the Court has consistently applied and interpreted its powers in a restrictive manner.³⁷⁵ For instance, when Roberts found that the success rate in appeals raising only ‘factual issues’ was low, she concluded that this was evidence of the Court acting restrictively.³⁷⁶ This ‘restrictive’ interpretation of the Court’s powers led to the observation that appellants who can point to a substantive significant error of law are most likely to be successful; and appellants who cannot are unlikely to be successful.³⁷⁷ Chapter 4 explains that this has become part of the meaning of the ‘unsafety test’ and part of the law which governs appeals.

In this thesis, a number of variables which could constitute ‘institutional’ factors are considered. In particular, there is a variable designed to capture the law relating to appeals. This measurement of the law embodies how judges perceive their roles, and how the ‘unsafety test’ has been interpreted. Other variables which capture the institutional position of the Court of Appeal, include whether permission to appeal was granted by the single judge or by the Full Court, and whether the appeal was by way of a *Criminal Case Review Commission* (CCRC) referral.

Utilising the desire to consider broader factors which could influence the decision-making of judges, a range of other models / theories have developed in recent years. Resnik developed a managerial model of judicial decision-making, which postulates that a primary concern of judges is to control their workload and calendars.³⁷⁸ She argued that managerial judging is problematic because it

³⁷⁵ See R Pattenden, *English Criminal Appeals 1844-1994* (Clarendon Press 1996); R Nobles and D Schiff, *Understanding Miscarriages of Justice: The Law, The Media, and the Inevitability of Crisis* (Oxford University Press 2000); S Roberts, ‘The Royal Commission on Criminal Justice and Factual Innocence: Remedying Wrongful Convictions in the Court of Appeal’ (2004) (1)(2) JJ 86; M Naughton (ed), *The Criminal Cases Review Commission: Hope for the Innocent?* (Palgrave Macmillan 2009); SJ Heaton, *A Critical Evaluation of Using Innocence as a Criterion in the Post-Conviction Process* (DPhil Thesis, University of East Anglia 2013).

³⁷⁶ *ibid.*

³⁷⁷ See S Roberts, ‘The Royal Commission on Criminal Justice and Factual Innocence: Remedying Wrongful Convictions in the Court of Appeal’ (2004) (1)(2) JJ 86.

³⁷⁸ J Resnik, ‘Managerial Judges’ (1983) 67 Harv L Rev 374.

appears to elevate speed over deliberation and impartiality.³⁷⁹ George and Yoon have recently explored the managerial model, and found stronger evidence for managerialism than for attitudinalism.³⁸⁰ Echoes of managerialism appear in some outputs from the Court of Appeal. In the 2015 / 16 Annual Report Hallett LJ, Vice President of the Criminal Division, demonstrated concerns with the management of the Court. She referred to the Court being 'burdened' with overly lengthy grounds of appeal, and expressed concern regarding the number of judges required for certain appeals.³⁸¹ She expressed similar concerns in the previous year's annual report.³⁸²

Managerial concerns may also be relevant in the variable which has been designed to capture the law applying to appeals. As will be discussed further in Chapter 4, the appeals process appears to favour appeals which raise procedural irregularities. As Spencer noted, the leave process in appeals appears well suited to locating procedural errors in trials, but it is less well suited to dealing with claims that the jury simply reached the wrong decision.³⁸³ Carefully reviewing the facts of the case, and receiving fresh evidence, take time and expend energy, which from a managerial perspective is difficult for the Court given its limited resources. This, ultimately, may feed into the 'unsafety test' which, like the leave process, appears to favour procedural irregularity appeals.

Other researchers have sought to relate judicial decision-making to general theories of human decision-making. As was discussed in Chapter 2, some researchers have sought to uncover whether behavioural heuristics and biases influence decision-making.³⁸⁴ Others have focussed upon the motivations behind human decision-making, and how this relates to judicial decision-making. For

³⁷⁹ *ibid*, 425.

³⁸⁰ TE George and AH Yoon, 'Chief Judges: The Limits of Attitudinal Theory and Possible Paradox of Managerial Judging' (2008) 61 (1) *Vand L Rev* 1.

³⁸¹ Court of Appeal, *In the Court of Appeal Criminal Division 2015 / 16* (Court of Appeal 2017) 4.

³⁸² Court of Appeal, *In the Court of Appeal Criminal Division 2014 / 15* (Court of Appeal 2016) 3.

³⁸³ JR Spencer, 'Does Our Present Criminal Appeal System Make Sense?' (2006) *Crim LR* 677, 684.

³⁸⁴ See, for instance AJ Wistrich, C Guthrie and JJ Rachlinski, 'Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding' (2005) 4 *U Penn L Rev* 1251.

instance, Baum argued that an inherent human trait – the desire to be liked and respected by people who are perceived as important to them – must also apply to judges as they, too, are human beings.³⁸⁵ Baum argued that this desire leads to changes in behaviour, and this too may influence the decisions which judges reach.³⁸⁶ A psychological analysis of the motivations of judges is beyond the design and scope of this study. It is emphasised that the study of judicial behaviour is now well developed, and in some countries, has moved beyond correlational study of the outcomes of cases.

A particular area of development relevant to this study is consideration of the decision-making of panels of judges.³⁸⁷ ‘Panel effects’ refers to how the members of multi-member courts might influence each other, and the decisions that a court reaches.³⁸⁸ The Court of Appeal sit in panels of at least three judges, and social psychological research has explored how small groups reach decisions.³⁸⁹ In American studies, the study of panel effects has focussed upon what happens when the panel is a mix of Democrat and Republican judges. Revesz found that judges were more likely to vote ideologically when they were sitting in panels with other judges from the same party.³⁹⁰ Furthermore, he found that the ideology of a judge’s colleagues was a better predictor of that judge’s vote than that judge’s own ideology.³⁹¹ Cross and Tiller similarly found that the ideological composition of panels was related to how judges vote.³⁹² They coupled the analysis of the ideological composition of panels with an analysis of legal doctrine. They argued that the presence of a political minority (particularly, a Democrat judge) within a

³⁸⁵ L Baum, *Judges and their Audiences: A Perspective on Judicial Behavior* (Princeton University Press 2008) 25.

³⁸⁶ *ibid.*

³⁸⁷ See KM Quinn ‘The Academic Study of Decision Making on Multimember Courts’ (2012) 100 Cal L Rev 1493.

³⁸⁸ *ibid.*, 1497.

³⁸⁹ See WL Martinek, ‘Judges as Members of Small Groups’ in D Klein, G Mitchell (eds) *The Psychology of Judicial Decision Making* (Oxford University Press 2010)

³⁹⁰ R Revesz, ‘Environmental Regulation, Ideology, and the DC Circuit’ (1997) 83 Va L Rev 1717.

³⁹¹ *ibid.*, 1764.

³⁹² FB Cross and EH Tiller, ‘Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing in the Federal Courts of Appeals’ (1998) 107Yale LJ 2155.

judicial panel moderated the panel when the majority were intending to ‘disobey’ legal doctrine.³⁹³

Not all studies have found panel effects. Sunstein, Schkade, and Ellman found no evidence of panel effects in several areas of law, including criminal appeals.³⁹⁴ Moreover, even when there appeared to be panel effects they were not large.³⁹⁵ The authors suggested that the minimal panel effect was due to different court compositions dampening the ideology of panels; the ‘disciplining effect of precedent and law’; and the dampening effect of legal culture.³⁹⁶ Kim has also studied panel effects. She sought to determine whether judges in politically mixed panels act strategically to avoid being overruled on appeal.³⁹⁷ She postulated that strategic judges would mediate their decisions to seek to prevent dissent from a minority judge.³⁹⁸ She found mixed evidence for this strategic decision-making. She found that the US Courts of Appeals panels did not seek to avoid review by the Supreme Court, but they did seek to avoid review by other Courts of Appeals panels.³⁹⁹ She suggested managerial reasons for this (further hearings by the Court meant more work for the Court),⁴⁰⁰ and reasons similar to those proposed by Baum: that the opinions of their Courts of Appeals colleagues were more important and salient to them.⁴⁰¹

It is difficult to properly assess panel effects in this study. This is because there is only one judgment per case in the Criminal Division of the Court of Appeal, and so it is not possible to observe what the role of each judge was in a particular judgment. This, it must be noted, means an important part of the decision-making

³⁹³ *ibid*, 2173.

³⁹⁴ CR Sunstein, D Schkade and LM Ellman, ‘Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation (2004) 90 Va L Rev 301, 334. Note that some of Sunstein’s other work did find evidence of panel effects, see CR Sunstein, *Why Societies Need Dissent* (Harvard University Press 2005).

³⁹⁵ *ibid*, 336.

³⁹⁶ *ibid*, 336-7.

³⁹⁷ PT Kim, ‘Deliberation and Strategy on the United States Courts of Appeals: An Empirical Exploration of Panel Effects’ (2008) 157 U Pa L Rev 1319.

³⁹⁸ *ibid*, 1345-6.

³⁹⁹ *ibid*, 1368-9.

⁴⁰⁰ *ibid*, 1369.

⁴⁰¹ *ibid*.

process – collegiality and deliberation – is omitted. One aspect of panel decision-making which can be explored, however, is the constitution of particular panels. Variables have been collected for the ranks of judges, which will help to determine whether judges in leadership roles (such as the Lord Chief Justice and the Vice President of the Criminal Division) are associated with deciding appeals in particular ways. This offers an opportunity to explore whether the dynamics of Court composition has any apparent role in decision-making.

3.7 The relationship between this thesis and previous judicial behaviour studies

As has been seen, in some countries, particularly the United States, scholars have been studying the decision-making of judges for decades. This has led to the development of the models of judicial decision-making which have been discussed above. There are, however, legitimate reasons to doubt whether such models can fully account for the complexities of judicial decision-making. The limitations of these kinds of studies are considered in depth throughout Chapters 5 and 6 of this thesis. There are considerable reasons to question the behavioural / attitudinal models in particular.⁴⁰² Furthermore, the relative absence of judicial studies in Britain means that the models have not been comprehensively tested in Britain. Accordingly, there is no real *a priori* empirical justification for hypothesising that such models would be successful in explaining judicial behaviour in Britain.

The research conducted in this thesis, therefore, does not seek to confirm whether any of these models appear to account for the decisions of judges. Rather, an explorative approach has been adopted. This study may be best understood as an explorative data analysis of a large amount of data collected

⁴⁰² This is particularly true of studies based upon the US Supreme Court Judicial Database, see See TE Pettys, 'Free Expression, In-Group Bias, and the Court's Conservatives: A Critique of the Epstein-Parker-Segal study' (2015) 63 Buff L Rev 1, and the discussion this critique generated: CL Boyd 'In defence of Empirical Legal Studies' (2015) 64 Buff L Rev 363; TM Hagle 'Two worlds, neither perfect: A Comment on the tension between legal and empirical legal studies' (2015) 63 Buff L Rev 379; C Shapiro 'Numbers, motivated reasoning, and empirical legal scholarship' (2015) 63 Buff L Rev 385.

from Court of Appeal judgments. In this analysis, there is an emphasis on the observation of whether patterns occur in the data.⁴⁰³ An explorative inductive approach is useful when the background reading on the topic 'did not lead to an expectation that results should come out in a particular way'.⁴⁰⁴

The relationship between this thesis and the field of judicial studies is to employ the methods developed by primarily American researchers, and apply those methods afresh in the Court of Appeal. This has a number of advantages. The study is not burdened by seeking to test whether one particular model explains decision-making. Furthermore, by analysing a range of variables which may to varying degrees be indicative of one or more model, it is possible to determine, in a tentative way, whether any model has sufficient traction for further analysis in the future. The history of judicial studies in the United States shows that simplistic behavioural models were replaced by more advanced attitudinal versions, which have since been continuously critiqued and questioned. This thesis has the advantage that these critiques have already taken place, and so this study can take account of the methodological difficulties which have been uncovered.

The variables which have been collected and analysed have been designed to capture the principle of impartiality. An exploration of impartiality is somewhat more general than an exploration of whether particular models explain judicial decision-making. This is a further benefit of the explorative inductive approach taken in this thesis. By focussing upon impartiality rather than a particular model, the study transcends attachment to any model. Rather, the variables are used to determine whether particular variables are associated with particular outcomes, to allow a discussion of what that means for the impartiality of the Court. Lastly, it should be remembered that it cannot be claimed that the variables completely

⁴⁰³ AT Jebb, S Parrigon and SE Woo, 'Exploratory Data Analysis as a Foundation of Inductive Research' (2017) HR Man Rev 265.

⁴⁰⁴ PE Spector et al, 'Moving the pendulum back to the middle: Reflections on and introduction to the inductive research special issue of Journal of Business and Psychology' (2014) 29 J Bus Psychol 499, 502.

capture the principle of impartiality. As such, the results of this study cannot be conclusive of partial or impartial decision-making, but the emergence of any patterns in the data can provide further points to consider regarding the Court's impartiality.

3.8 Conclusion

It has been explained that the principle of conducting empirical analysis on the decision-making of judges can claim a heritage from the American Legal Realists. The work conducted in this thesis is tied to the Empirical Legal Studies (ELS) research movement. The majority of ELS work on the decision-making of judges is American, and this is likely to be due to the influence of American Legal Realism on the American legal academy, an influence which is much more muted in other jurisdictions. As a result, quantitative, positivistic, analyses of the decision-making of judges, utilising more complex social science tools and methods, are relatively rare in Britain, and unique for the Court of Appeal (Criminal Division). This thesis has therefore observed this gap in research and literature and sought to complete it with an analysis of decision-making in the England and Wales Court of Appeal (Criminal Division).

The extensive research programmes in the United States has led to the development, and continued critique and refinement of judicial studies. There has been proposed a number of 'models' of judicial decision-making, which have been empirically tested for several decades in the United States. These include the behavioural model, attitudinal model, strategic model, managerial, and psychological models. Analysis of panel effects is becoming increasingly important. This thesis has utilised this literature in order to commence this explorative study of the Court.

The next chapter analyses the powers of the Court of Appeal and the background to the 'unsafety test'. It will be shown that the 'unsafety test' developed gradually in an attempt to conduce a more liberal attitude in the Court. However, this was

unlikely to work in practice, because to conduce a more liberal approach the powers of the Court simply became more discretionary. As a result of this discretion, as explained in Chapter 2, there is a possibility of partiality. This makes the Court of Appeal a good candidate for empirical analysis.

Chapter 4

The England and Wales Court of Appeal (Criminal Division)

Introduction

This chapter begins by analysing the purposes of the Court of Appeal. It has a dual purpose in attempting to correct miscarriages of justice by quashing the convictions of people who may be innocent, and upholding due process by quashing convictions achieved following procedural errors or breaches of due process. The Court of Appeal's review function serves an important legitimising purpose, in that it operates as a review of whether the convictions it examines have 'moral authority'. Moral authority pertains to the legitimacy of guilty verdicts, in that it should be a morally justified statement of the defendant's blameworthiness and fitness for punishment.⁴⁰⁵ Moral authority will derive from the conviction's factual accuracy, and its respect for the moral and political principles which underlie the criminal process.⁴⁰⁶ Given this function of the Court in reviewing the legitimacy of convictions, it is important to know whether the Court itself has legitimacy to render decisions.

All of the cases analysed in this thesis were decided while the Court was operating the 'unsafety test'. Thus, it is important to understand how the test works, and this chapter provides a thorough analysis of the test. This chapter briefly analyses the history of the Court's powers. It will be shown that the Court's powers have progressively become looser and more discretionary in nature, and 'open-textured'. The 'unsafety test' appears to provide the Court with maximum flexibility and discretion. However, this flexibility is dampened to an extent by the Court having stipulated what is necessary for a conviction to be unsafe. It is only if an appellant can meet the tests built into the 'unsafety test' by case law, that the Court can quash a conviction. Thus, it could be said that while the 'unsafety test' appears discretionary, it is often guided by legal rules and principles.

⁴⁰⁵ I Dennis, 'The Right to Confront Witnesses: Meanings, Myths and Human Rights' (2010) 4 Crim LR 255, 259-60.

⁴⁰⁶ *ibid.*

4.1 Purposes and function of the Court of Appeal

The purpose of having an appeal system is inextricably linked to the general overall purpose, function, or aim, of criminal justice. Rule 1 of the *Criminal Procedure Rules*⁴⁰⁷ provides an ‘overriding objective’ that criminal cases are dealt with ‘justly’. Dealing with cases justly includes a number of, often conflicting, principles or subsidiary aims. The first three are most important for present purposes, and require, firstly, ensuring that the innocent are acquitted and the guilty are convicted.⁴⁰⁸ Secondly, the prosecution and the defence must be treated fairly.⁴⁰⁹ Thirdly, a defendant’s rights under Article 6 of the *European Convention on Human Rights* should be recognised.⁴¹⁰

There are some difficulties with these aims which should be mentioned. Firstly, the aim of ‘convicting the guilty’ may appear somewhat incongruous with the presumption of innocence. Since all persons on trial are presumed to be innocent, does it make sense that the supposed aim of the trial is to ‘convict the guilty’, given that a person is only legally ‘guilty’ once the conviction has occurred? The aim is suggestive of guilt being a condition which exists ‘out there’, needing only to be discovered by the trial. At an abstract level this might be true, but as Nobles and Schiff argued, it is doubtful whether guilt is a thing which exists ‘out there’, given that guilt or otherwise often concerns contested versions of past events, coupled with ‘legal guilt’.⁴¹¹

Some of the aims appear to be in conflict with each other. Even appellants who are in an abstract sense ‘guilty’, or those believed to be guilty, are entitled to have their due process rights recognised. This raises the question of what happens when they have their rights abused? Moreover, is it possible to ‘ensure’ the acquittal of the innocent, whilst simultaneously attempting to do justice by

⁴⁰⁷ Available at <https://www.justice.gov.uk/courts/procedure-rules/criminal/rulesmenu-2015> <accessed 23 August 2016.

⁴⁰⁸ See rule 1.1.2(a).

⁴⁰⁹ *ibid*, rule 1.1.2(b).

⁴¹⁰ *ibid*, rule 1.1.2(c).

⁴¹¹ See R Nobles and D Schiff, *Understanding Miscarriages of Justice: The Law, The Media, and the Inevitability of Crisis* (Oxford University Press 2000) 21.

punishing wrongdoers? The epistemological limits of the criminal trial, in that belief 'beyond reasonable doubt' is required, and not 'belief beyond all certainty', means that error is always possible. Is there an acceptable error rate?⁴¹² Or, is it not the case that only way to 'ensure' the innocent are always acquitted would be to not convict anyone?⁴¹³ Moreover, can any of these principles be achieved whilst at the same time conforming to principle (e) of Rule 1 of the *Criminal Procedure Rules*: dealing with cases efficiently and expeditiously? With unlimited time and resources the aims might be possible, but that is not the reality for the criminal justice system generally, and the Court of Appeal in particular. Such questions go to core of the role of the Court's institutional position in reviewing convictions and quashing those which it thinks are unsafe.

Spencer argued that the Court of Appeal is in place to remedy two distinct ills.⁴¹⁴ The first is to remedy a miscarriage of justice in the sense of the conviction of people who are innocent; the second is to remedy a failure of due process of law, to an extent, irrespective of actual innocence or guilt.⁴¹⁵ These accord with the general overriding objective within the *Rules* to do justice. Similarly, Dennis argued that the fundamental function of the Court is to review the legitimacy of convictions.⁴¹⁶ This requires a review of whether the conviction is factually accurate, carries sufficient 'moral authority' to justify continued punishment, and is 'founded on the rule of law'.⁴¹⁷ Convictions can only be safe if the Court is satisfied of these requirements. The Court's role is to do justice, because convictions failing these test could not be just. 'Morally authoritative' and 'founded on the rule of law' are related to the circumstances in which breaches of fairness or other procedures can render a conviction unsafe even if they do not cast doubt on the accuracy of the verdict.

⁴¹² Most famously, Blackstone offered the 10:1 ratio in his *Commentaries on the Laws on England*, (1809) vol 4, 358. It is not necessary to discuss the question of the appropriate 'ratio' further.

⁴¹³ See V Halvorsen, 'Is It Better That Ten Guilty Persons Go Free Than One Innocent Person Be Convicted?' (2004) 23 *Crim Just Ethics* 3.

⁴¹⁴ JR Spencer 'Does Our Present Criminal Appeal System Make Sense?' (2006) *Crim LR* 677.

⁴¹⁵ *ibid*, 683.

⁴¹⁶ I Dennis, 'Fair Trials and Safe Convictions' [2003] *CLP* 211.

⁴¹⁷ *ibid*, 236.

Dennis argued that the Court has adopted a clear position in which a conviction will always be unsafe if the Court thinks that there is a reasonable doubt about the conviction's factual accuracy.⁴¹⁸ Furthermore, substantial breaches of the rule of law will always cause convictions to be unsafe, if it was such that it was unfair to try the appellant at all.⁴¹⁹ He also suggested that in cases where it is alleged that breaches of procedure occurred, the essential question is whether the breach prejudiced the appellant to the extent that the outcome of the trial might have been different but for the irregularity.⁴²⁰ Dennis's arguments will be analysed below, when the 'unsafety test' is considered in greater depth.

The complexities of the functions of the Court are well illuminated by the case of *R v Paris and others*,⁴²¹ also commonly known as the *Cardiff Three* case. The three appellants were convicted of the murder of Lynette White. One appellant, Miller, who was mentally impaired and had a reading age of 11, confessed to the murder of White (who was his girlfriend). The others had witnesses testify against them placing them at the scene. The investigation occurred after the enactment of the *Police and Criminal Evidence Act 1984 (PACE)*, which provides that confessions gained by oppressive questioning are inadmissible.⁴²² The appellants challenged the admissibility of the confessions at trial, alleging oppression, but the confessions were ruled admissible by the trial judge. Two years later, the convictions were quashed. In the appeal, the Court of Appeal listened to the tape recordings of the police interviews, parts of which had not been made available to the trial judge, and concluded that Miller had been 'bullied and hectorred' by the police into confessing.⁴²³ He was asked the same question over 300 hundred times, in order to induce confession.

The *Cardiff Three* could be considered a classic example of the Court of Appeal upholding due process values in quashing the convictions due to procedural

⁴¹⁸ *ibid*, 219.

⁴¹⁹ *ibid*, 214.

⁴²⁰ *ibid* 230.

⁴²¹ (1993) 97 Cr App R 99.

⁴²² *Police and Criminal Evidence Act (1984) (PACE)* s 76.

⁴²³ *Paris*, (n 421) 103.

irregularities and misbehaviour by the authorities. This would be such a classic example, except that it is now known, as far as it can be, that as well as having their due process rights abused, they were also innocent of the murder. The real killer was convicted in 2003 and absolved them of any involvement.⁴²⁴ The witnesses against the Three were convicted of perjury and sentenced to imprisonment.⁴²⁵ The judge sentencing the perjurers noted that they were 'seriously hounded, bullied, threatened, abused, and manipulated by the police' into making false statements against the *Cardiff Three*.⁴²⁶ A criminal case against the police officers concerned collapsed in 2011, after prosecution evidence against them disappeared. The evidence was found a short time later in a cupboard.⁴²⁷

The fact that the *Cardiff Three* are known to have also been innocent serves to obscure that the convictions were quashed not because they were innocent, but because the Court of Appeal found that the confession should not have been admitted at trial. Naughton is critical for the Court because of this. He stated that the Court did not quash the convictions because the Court wanted to correct an apparent wrong, the conviction of innocent people, but because of this irregularity.⁴²⁸ He said that this is evidence that the Court is not concerned with correcting miscarriages of justice as understood from a 'lay' perspective, but only in a legalistic sense.⁴²⁹ This criticism of the Court does not seem well placed. The Court could not have quashed the conviction on the basis of their innocence because, at the time the convictions were quashed, the Court could not have known that they were actually innocent. It may even be said that this case demonstrates the Court taking a strongly pro-active approach in seeking to correct miscarriages of justice because the Court quashed the conviction *despite*

⁴²⁴ See S Morris 'Cardiff Three Are Completely Innocent, Says Convicted Killer of Lynette White' (The Guardian, 13 July 2011).

⁴²⁵ See BBC News Online 'Three Jailed for Murder Case Lies' 19 December 2008 <http://news.bbc.co.uk/1/hi/wales/7792501.stm>. <18 August 2016>

⁴²⁶ *ibid.*

⁴²⁷ See Independent Police Complaints Commission, Destruction of Specific Documents Leading to the Collapse of the R v Mouncher & others trial, *Report*, (IPCC 2011).

⁴²⁸ M Naughton, *The Innocent and the Criminal Justice System: A Sociological Analysis of Miscarriages of Justice* (Palgrave Macmillan 2013) 154.

⁴²⁹ *ibid* 151 – 4.

there not being conclusive proof of innocence at the time of the appeal. The convictions were quashed because the breaches of rules were so obvious and egregious, and so clearly in flagrant disregard of PACE, that the convictions had to be quashed. They lacked any moral authority. Therefore the convictions were quashed not because innocence was proved, but because there was a risk that real injustice had been done. It is unclear how this demonstrates that the Court is not concerned with correcting the convictions of the innocent.⁴³⁰

Quashing convictions for procedural irregularities may at times appear difficult to justify. One example of a case which is difficult to justify is *R v Weir*.⁴³¹ His conviction for murder was quashed because the DNA evidence against him should have been destroyed and removed from the National DNA database, as required by section 64 of PACE. The Crown's appeal against the Court of Appeal's decision was dismissed by the House of Lords because the appeal was lodged one day late.⁴³² If the Crown's appeal had been heard by the House, it is very likely to have been successful because the decision in *Attorney General's Reference (No 3 of 1999)*⁴³³ held that failing to destroy DNA samples under section 64 did not automatically make the evidence inadmissible. Indeed, Lord Steyn implied that the Court of Appeal decision in *Weir* was 'absurd'⁴³⁴ and 'wrong'.⁴³⁵ It is therefore highly unlikely that the same decision would be reached now, but the case does demonstrate the conflict between protecting due process rights and ensuring the conviction of the likely guilty.

The overall purpose of the Court of Appeal is to do justice. Within that overall aim, however, there are a variety of conflicting principles and interests. The Court is in a position of having to balance the overriding objective principles of ensuring that innocent appellants have their convictions quashed, against dismissing unmeritorious appeals. Furthermore, the interest of fairness, which is afforded to

⁴³⁰ As Naughton alleged at *ibid* 142.

⁴³¹ *R v Weir* Times 16 June 2000.

⁴³² *R v Weir* [2001] 1 WLR 421.

⁴³³ [2001] 1 AC 91.

⁴³⁴ *ibid*, 118.

⁴³⁵ *ibid*, 120.

both guilty and innocent appellants, is balanced against the interest of upholding the conviction of the guilty. All these interests must then be balanced against maintaining the integrity of the system, in particular the finality of jury verdicts and the system of trial by jury in which it is the jury tasked with determining guilt or innocence in Crown Court trials. There has been concern that the Court is balanced too far in the direction of upholding the integrity of the jury system, and finality, rather than doing justice in the sense of quashing the conviction of the innocent. These concerns led ultimately to the ‘unsafety test’ under which the Court currently operates. These issues, and the ‘unsafety test’, are returned to below.

4.2 Criminal justice system statistics

The most recent available complete data relating to criminal appeals is for the year 2016.⁴³⁶ The appeals studied in this thesis, however, are from the time period 2006 to 2010. Data is available for this time period from a number of different sources.⁴³⁷ In 2016, 1.5 million cases were received by the magistrates’ court system.⁴³⁸ Of these, 326000 were either-way offences; 32000 were indictable only offences; 562000 were summary motoring offences; 539000 were summary offences; and 68000 were for breaches of orders.⁴³⁹ In 2010, 1.7 million cases were received by the magistrates’ court.⁴⁴⁰ The combined number of indictable and triable either-way offences was 410000; 590000 were summary motoring offences; 546000 were summary only offences; 117000 were breaches; and 131000 were youth proceedings.⁴⁴¹

⁴³⁶ All statistics are taken from the tables contained in *Criminal Court Statistics: October to December 2016 (main tables)* available from <https://www.gov.uk/government/statistics/criminal-court-statistics-quarterly-october-to-december-2016> (accessed 19/06/2017).

⁴³⁷ See *Court Statistics Quarterly January – March 2012 main tables*. Available from <https://www.gov.uk/government/statistics/court-statistics-quarterly-earlier-editions-in-the-series> (accessed 19/06/2017).

⁴³⁸ 2016 data, table M1.

⁴³⁹ *ibid.*

⁴⁴⁰ 2010 data, table 3.1.

⁴⁴¹ *ibid.*

In 2016, 116000 cases were received by the Crown Court system.⁴⁴² 46000 related to either-way offences; 27000 were indictable only offences; 31000 were committals for sentence; and 10000 were appeals from magistrates' court decisions.⁴⁴³ In 2010 the Crown Court received 152000 cases, of which 63000 were triable either-way offences; 34000 were indictable only offences; 41000 were committals for sentence; and there were 13000 appeals from magistrates' courts.⁴⁴⁴

The most recent available full year statistics on conviction rates in the Crown Court are from 2015.⁴⁴⁵ These show that in 2015 93000 individual defendants were tried (109000 in 2010).⁴⁴⁶ Of these, 64000 pleaded guilty to all counts (76000 in 2010). Of those who pleaded not guilty in the Crown Court, 12000 individuals were convicted in 2015. The following Table shows the number of individual defendants convicted in a Crown Court following a not guilty plea in the period 2007 to 2015, thus encompassing part of the period under analysis in this thesis.⁴⁴⁷

⁴⁴² 2016 data, table C1.

⁴⁴³ *ibid.*

⁴⁴⁴ *ibid.*

⁴⁴⁵ See *Criminal Court Statistics Quarterly January to March 2016*.

⁴⁴⁶ *ibid.*, table AC6.

⁴⁴⁷ Produced from *ibid.*

Table 4.1: Number of defendants convicted in a Crown Court following not guilty plea 2007-2015.

Year	Number
2007	11073
2008	11137
2009	11252
2010	11957
2011	12275
2011	12490
2013	11518
2014	11121
2015	12106

It will be seen that the number has remained fairly static. Out of a total of more than 1.5 million criminal cases, therefore, only a small proportion of defendants are convicted of the most serious crimes in the Crown Court following a not guilty plea. As will be discussed further below, only a small proportion of these will apply for permission to appeal their convictions.

There is further data available for the number of persons convicted of the offence of rape between 2006 and 2010. In 2013, the Home Office produced a statistical bulletin on sexual offending.⁴⁴⁸ The statistics break down the offence of rape into several categories. Firstly, rape is divided into rape of a male and rape of a female.⁴⁴⁹ Secondly, there is rape of a female over the age of 16 charged as rape under section 1 of the *Sexual Offences Act 2003*, and rape of a child under the age of 13, (charged as rape under section 5). The statistics also show section 1 rape of a child aged under 16 (but presumably 13 or over). There is no separate offence of rape for this category; it would be charged under section 1.⁴⁵⁰ These same categories are provided for rape of a male. The data also shows attempts

⁴⁴⁸ *An Overview of Sexual Offending in England and Wales* (Home Office, Ministry of Justice 2013).

⁴⁴⁹ *ibid*, Sexual Offending Overview Tables, table 4.1.

⁴⁵⁰ *ibid*, see glossary.

of these offences. The following Table reproduces the data relating to the number of persons convicted of these kinds of rape against a female.⁴⁵¹ As they are not considered in this thesis, attempts are omitted. This Table includes persons who either pleaded guilty or who were convicted.

Table 4.2: Persons convicted of rape of a female 2006 - 2010.

	2006	2007	2008	2009	2010
Under 13	85	91	141	143	167
13 – under 16	219	236	241	273	271
16 or over	384	375	382	413	445
Total	688	702	764	829	883

The following Table shows the same data for persons convicted of these categories of rape against a male. The total number of persons convicted of an offence of rape between 2006 and 2010 is shown in Table 4.4. These Tables include persons who either pleaded guilty or who were convicted.

Table 4.3: Persons convicted of rape of a male 2006 - 2010.

	2006	2007	2008	2009	2010
Under 13	32	37	33	28	46
13 – under 16	15	23	14	15	17
16 or over	20	16	12	14	14
Total	67	76	59	57	77

Table 4.4: Total number of persons convicted of rape 2006 – 2010.

Year	2006	2007	2007	2009	2010
Number	755	778	823	886	960

⁴⁵¹ Reproduced from *ibid*, table 4.1.

This thesis has included the age of the complainant in a rape appeal and the deceased in a murder appeal as independent variables. These above age categories are used as the ages for the independent variables. There is an additional variable for those aged 16 and 17, and so who are still considered children. This has the benefit that the variables used for rape are mapped onto the *Sexual Offences Act 2003*. However, this does mean that these age categories, specifically designed for rape, are imposed onto murder appeals. The specific decisions relating to the coding of the murder and rape appeals are returned to and analysed in Chapter 6.

Statistics are also available for the number of persons convicted of murder between 2006 and 2010.⁴⁵² The following Table reproduces this data. This Table includes persons who either pleaded guilty or who were convicted.

Table 4.5: Number of persons convicted of murder 2006 – 2010.

Year	Number
2006	358
2007	315
2008	328
2009	267
2010	141

Those who have been convicted in a Crown Court have an entitlement to seek 'leave' (permission) from the Court of Appeal to appeal against the conviction. As was shown above, approximately 11000 - 12000 persons are convicted each year in a Crown Court following a not guilty plea. Whilst defendants who pleaded guilty can, and do, occasionally appeal their convictions, it is relatively rare for those who pleaded guilty to do so. As is discussed below, appeals following a guilty plea are unlikely to be successful.

⁴⁵² Data collected from *Homicides, Firearm Offences, and Intimate Violence 2009-10 Supplementary volume 2* (Home Office 2011) table 1.10.

An appellant who wishes to appeal must lodge an application to the Court which will initially be decided by a single High Court judge. The single judge can either grant leave, or refuse leave. If leave is refused, it can be renewed before a Full Court which will decide whether to grant leave. Table 4.6 shows the number of applications for permission to appeal against conviction received per year between 2006 and 2010; the number of applications granted by the single judge; the number of failed applications renewed; and the number of applications granted by the Full Court.⁴⁵³

Table 4.6: Applications received 2006 – 2010.

Year	Applications	Granted Single judge	Renewed Applications	Granted By Full Court
2006	1595	291	481	137
2007	1508	288	520	125
2007	1588	212	400	146
2009	1435	275	477	117
2010	1488	242	370	148

A number of observations can be made about this Table. Most applications are declined by the single judge – fewer than 20% of applications are granted permission by the single judge. Of the unsuccessful applicants who decide to renew their applications, the chances of being granted permission by the Full Court are slightly improved. There has also been a steadily declining number of applications received by the Court, and this continues to the present day. In the 12 months ending October 2016, the Court of Appeal received 1417 applications for permission to appeal against conviction.⁴⁵⁴

The method of permission being granted are independent variables in this thesis. In particular, whether the case is heard following permission by the single judge,

⁴⁵³ Data extracted from *Court Statistics Quarterly January to March 2014* table 5.7.

⁴⁵⁴ *In the Court of Appeal Criminal Division Annual report 2015-16* (2017) 27.

the Full Court, or the CCRC. This variable relates to the institutional position of the Court of Appeal, and how the processes of the Court function.

The above Table highlighted statistics relating to the permission to appeal process in the Court of Appeal. Table 1.1 in Chapter 1 compared the number of cases heard per year between 2006 and 2010 with the number of combined murder and rape appeals per year. This is reproduced here as Table 4.7.

Table 4.7: Comparison of overall Court of Appeal workload and murder and rape workload (2006-2010).

	2006	2007	2008	2009	2010	Totals
Sampled Appeals	120	91	66	91	104	472
Overall workload	572	523	438	430	496	2459
% workload in sample	20.9	17.3	15.0	21.1	20.9	19.1

Since 2010, the number of cases heard by the Court of Appeal has declined significantly. In 2016 the Court heard only 260 appeals against conviction.⁴⁵⁵ These data relating to the appeals heard between 2006 and 2010 are returned to in Chapter 7.

The criminal justice statistics recounted in this section reveal a number of points. The Court of Appeal hears only a small fraction of all criminal cases. If the 11000 - 12000 defendants convicted in the Crown Court following a not guilty plea in 2016 is the most appropriate the denominator, and the 260 appeals heard in 2016 is the numerator, the Court heard 0.02% of cases within its jurisdiction. The Court's role is much more significant qualitatively than quantitatively. This relates to, as explained in section 4.1, the role of the Court in ensuring that due process is adhered to, and its role in seeking to correct miscarriages of justice. Despite

⁴⁵⁵ *ibid*, 25.

its quantitatively small reach, the Court is a suitable body for empirical study owing to the importance of the Court qualitatively.

4.3 A short history of the Court's powers

The Court of Criminal Appeal (as the Court of Appeal (Criminal Division) was known until 1966),⁴⁵⁶ was created by the *Criminal Appeal Act 1907*. The Act afforded the Court the power to allow an appeal against conviction if it felt the conviction was, (section 4(a)): 'unreasonable or could not be supported having regard to the evidence,' or (b), or should be 'set aside on the ground of any wrong decision on a question of law', or, (c), there was a 'miscarriage of justice'. There was an additional proviso that, notwithstanding a ground was made out, the Court could dismiss the appeal if it was satisfied that 'no substantial miscarriage of justice had occurred'. Section 9 of the Act gave the Court wide discretion to, if thought to be in the interests of justice, order the production of any 'document, exhibit or any other thing' which appeared necessary to determine the case, as well as call any witnesses. This afforded the Court the power to admit fresh evidence upon the application of an appellant. Section 4(a) of the 1907 Act invited the Court to consider the factual accuracy of verdicts, and enter into the territory of the jury. Allowing appeals on questions of fact was an important part of the 1907 Act, as 'to say that the Court should hear arguments on points of law, and treat all findings of fact as conclusively established, would be to reduce the Court to futility'.⁴⁵⁷

It will be observed that the Court's powers under the 1907 Act were broad, yet fairly specific and explicit. The textual difference between these powers and the single ground known as the 'unsafety test' is stark. It may be said that section 4 of the 1907 Act was a somewhat more 'rule-based' provision, whilst 'unsafety' is more open-textured and difficult to define. In creating the Court of Criminal Appeal, it was envisaged that its role would be to correct 'all matters such as the misapprehension of the judges, and the misleading of the jury, that [make]

⁴⁵⁶ The Court of Criminal Appeal was abolished by the Criminal Appeal Act 1966, to be replaced by the Court of Appeal Criminal and Civil Divisions.

⁴⁵⁷ See Parl Deb HC 29 July 1907, vol 179, col 587-8.

criminal trials sometimes unsatisfactory'.⁴⁵⁸ Clearly, this envisaged some kind of critical scrutiny of the decisions of juries. The Court's powers recounted above, however, appeared ill-suited to this task. In a debate on the *Criminal Appeal Bill* in 1907, concern was raised by F.E. Smith, who was later to serve as Lord Chancellor, that the Court's powers would not serve its purpose.⁴⁵⁹ He noted that the Court's proposed powers were based on the powers of the Court of Appeal, which at the time heard only civil appeals.⁴⁶⁰ He discussed the case of *Metropolitan Railway Co. v Wright*,⁴⁶¹ in which the House of Lords considered the scope of review in civil cases. In that case, the Earl of Selborne said that to overrule a civil jury's decision there must be 'such a preponderance of evidence ... as to make it unreasonable, and almost perverse, that the jury when instructed and assisted properly by the judge should return such a verdict'.⁴⁶² If this is what the Court's powers were to be based upon, Smith argued, the Court's powers would be seriously constrained.⁴⁶³

The primary difficulty for the Court was section 4(a) which invited the Court to consider whether the verdict was unreasonable or not supported by the evidence. The controversy stemmed from under what circumstances the Court would, or could, decide a conviction was unreasonable or not supported by the evidence. The Court frequently made statements similar to that made by Earl Selborne in the civil context. In the first appeal against conviction which was heard by the Court, the Lord Chief Justice said: 'it must be understood that we are not here to retry the case where there was evidence proper to be left to the jury upon which they could come to the conclusion at which they had arrived'.⁴⁶⁴ In *R v McGrath*,⁴⁶⁵ Lord Goddard stated: 'where there is evidence on which a jury can act and there has been a proper direction to the jury this court cannot substitute

⁴⁵⁸ *ibid*, 588.

⁴⁵⁹ See Parl Deb HC 29 July 1907, vol 179, col 586-8

⁴⁶⁰ *ibid*, col 634.

⁴⁶¹ (1886) 11 App Cas 152.

⁴⁶² *ibid*, 153.

⁴⁶³ See n 459, above.

⁴⁶⁴ *R v Williamson* (1909) 1 Cr App R 3

⁴⁶⁵ [1949] 2 All ER 495.

itself for the jury and retry the case'.⁴⁶⁶ In an early review of the Court, Ross⁴⁶⁷ noted that 'so great has been the regard paid by the Court to "trial by jury" that cases are extremely rare in which the conviction has been quashed solely on the ground ... [that it] was unreasonable'.⁴⁶⁸ He did note, however, that there was no such reluctance to put themselves in the place of the jury when deciding what effect a wrong decision on a question of law would have had on the jury.⁴⁶⁹

By the 1950s concern began to grow that the Court was not operating as it should. Nobles and Schiff observe the 1950s as being the 'high watermark of judicial non-receptivity' in particular in relation to fresh evidence under section 9 of the 1907 Act.⁴⁷⁰ The only statutory question for the Court when asked to exercise its powers to receive fresh evidence was whether it was 'necessary or expedient in the interest of justice'. The Court created its own hurdles which an appellant had to surmount. The hurdles to the admission of evidence were developed gradually⁴⁷¹ and were summarised by Lord Parker in *R v Parks*⁴⁷² as being:

'First, the evidence that it is sought to call must be evidence which was not available at the trial. Secondly...it must be evidence relevant to the issues. Thirdly, it must be evidence which is credible evidence in the sense that it is well capable of belief; it is not for this court to decide whether it is to be believed or not, but evidence which is capable of belief. Fourthly, the court will...consider whether there might have been a reasonable doubt in the minds of the jury as to the guilt of the appellant if that evidence had been given together with the other evidence at the trial'.⁴⁷³

The operation of these rules, especially the rule requiring that the evidence was not available at trial, caused difficulties for appellants. Spencer says that the

⁴⁶⁶ *ibid*, 497.

⁴⁶⁷ RE Ross, *The Court of Criminal Appeal* (Butterworth & Co 1911).

⁴⁶⁸ *ibid*, 88.

⁴⁶⁹ *ibid*, 89-90.

⁴⁷⁰ See R Nobles and D Schiff, *Understanding Miscarriages of Justice: The Law, The Media, and the Inevitability of Crisis* (Oxford University Press 2000) 61.

⁴⁷¹ See R Pattenden, *English Criminal Appeals 1844-1994* (Clarendon Press 1996), 131.

⁴⁷² (1962) 46 Cr App R 29.

⁴⁷³ *ibid*, 32.

Court made its section 9 powers as narrow as possible by inventing a rule that fresh evidence was not fresh if it could, with due diligence, have been called by the defence at trial.⁴⁷⁴ Thus, fresh evidence which was missed due to incompetence was not fresh. This approach to fresh evidence signalled that the Court was generally unreceptive to appeals based on fresh evidence or claims that the jury had simply made a mistake. The Court would very frequently reiterate the exceptional nature of receiving fresh evidence, in order to avoid an appeal becoming a retrial.⁴⁷⁵

The operation of the Court may have been hindered by the lack of a retrial power. The first major governmental report into the operation of the Court was the *Tucker Committee*.⁴⁷⁶ It was convened to consider whether the Court should be given the power to order a retrial, a power which was absent from the original Act. It proposed by a majority that the Court should be granted the power to order a retrial when convictions were quashed on the basis of fresh evidence. It was hoped by the majority of the Committee that providing the option of a retrial would make the Court more receptive to receiving fresh evidence.⁴⁷⁷ Although the Committee reported in 1954, it was not until the *Criminal Appeal Act 1964* that the power to order a retrial was granted.⁴⁷⁸

In 1964 the Parliamentary group JUSTICE issued a highly critical review of the Court, stating that 'a very restricted view has been taken of the Court's power'.⁴⁷⁹ In 1965 the *Donovan Committee* issued its report into the constitution and decision-making of the Court.⁴⁸⁰ The Committee's concerns stemmed from what

⁴⁷⁴ JR Spencer 'Criminal Law and Criminal Appeals: The Tail that Wags the Dog' [1982] Crim LR 260, 264.

⁴⁷⁵ For instance *R v Brown* (1910) 4 Cr App R 104; *R v Tellett* (1921) 15 Cr App R 159; *R v Mason* (1924) 17 Cr App R 160; See especially *R v Rowland* (1948) 32 Cr App R 29. Numerous other cases could just as easily have been cited.

⁴⁷⁶ Departmental Committee on New Trials in Criminal Cases, *Report*, [Tucker Committee] (Cmd. 9150, 1954).

⁴⁷⁷ *ibid*, [25]

⁴⁷⁸ Criminal Appeal Act 1964, s 1.

⁴⁷⁹ JUSTICE Committee, *Criminal Appeals* (Stevens and Sons 1964) [59].

⁴⁸⁰ Interdepartmental Committee on the Court of Criminal Appeal, *Report*, [The Donovan Committee] (Cmd 2755, 1965).

it thought was a defect with the drafting of the 1907 Act. It thought that the powers under section 4(a)-(c) of the 1907 Act overlapped and may have conflicted with each other. As the Committee noted:

‘If there was credible evidence both ways, and the jury accepted evidence pointing towards guilt, it is difficult to say that the verdict was ‘unreasonable’ or could not ‘be supported having regard to the evidence’ or that there was a ‘miscarriage of justice’.⁴⁸¹

The Committee recommended reformulating the Court’s powers. It also recommended that the Court of Criminal Appeal be abolished and the ‘Court of Appeal’ be split into Civil and Criminal Divisions. Both changes were enacted by the *Criminal Appeal Act 1966*, and the changes, including the right to order a retrial, were consolidated by the *Criminal Appeal Act 1968*. The powers of the Court became that it was to allow an appeal if it thought the conviction was: section 2:

- (a) unsafe or unsatisfactory; or
- (b) the judgment of the court of trial should be set aside on the ground of a wrong decision of any question of law; or
- (c) there was a material irregularity in the course of the trial.

The ‘unsafe or unsatisfactory’ test was originally proposed as an amendment by F.E. Smith in the debate to the 1907 *Bill*, referred to above.⁴⁸² The Attorney General rejected the amendment, due to his concern that the proposed amendment was ‘loose to the point of obscurity and of being unscientific, inasmuch as they would have no precision’.⁴⁸³ Nevertheless, sixty years later the *Donovan Committee* adopted Smith’s proposed wording.

⁴⁸¹ *ibid*, [141].

⁴⁸² HC Deb 29 July 1907, vol 179, col 634

⁴⁸³ *ibid*, col 638.

The *Donovan Committee* thought the new words were broader than the original formulation because it suspected that the new wording would lead to more convicted people applying to the Court, as they would 'see new hope in the new provision'.⁴⁸⁴ It thought that even under the 1907 Act the Court occasionally 'acted as a jury and come to the conclusion that on the totality of the evidence ... it would be unsafe to allow the verdict of guilty to stand'.⁴⁸⁵ This was despite, under the 1907 Act, it being doubtful whether the Court had that power. The Committee thought that the 'unsafe or unsatisfactory' test would remove any doubt that the role of the Court is to do justice, 'which includes the avoidance of possible injustice'.⁴⁸⁶

The proviso allowing the Court to uphold a conviction if it was sure that there had been no 'substantial miscarriage of justice' was retained, subject to the deletion of the word 'substantial'. The Committee noted that there had been criticism relating to how the Court applied its powers to receive fresh evidence.⁴⁸⁷ Section 9 of the 1907 Act became section 23 of the 1968 Act, which now provided that the Court had the power to receive fresh evidence if they thought it necessary or expedient in the interests of justice to do so. This was coupled with the duty to admit evidence which was credible and relevant and there was an explanation for not adducing it at trial, unless it would not afford any ground for allowing the appeal. The duty to admit fresh evidence if it was relevant and credible was recommended to conduce to Court to act in a way which would ensure 'so far as possible that any miscarriages of justice will be avoided or corrected'.⁴⁸⁸

There were two significant decisions shortly following the 1968 Act: *R v Cooper*,⁴⁸⁹ and *Stafford v DPP*.⁴⁹⁰ Lord Widgery in *R v Cooper* developed the doctrine of what is known as 'lurking doubt'. He stated that the new 'unsafe or

⁴⁸⁴ Donovan Committee (n 480) at [150].

⁴⁸⁵ *ibid*, [148].

⁴⁸⁶ *ibid*, [149].

⁴⁸⁷ *ibid*, [133].

⁴⁸⁸ *ibid*, [136].

⁴⁸⁹ (1969) 53 Cr App R 82.

⁴⁹⁰ [1974] AC 878.

unsatisfactory' ground meant something different to the 'unreasonable or not supported by the evidence' ground found in the 1907 Act. The new test meant that if all the evidence had been before the jury, and the evidence was correctly summed up, the Court could still quash the conviction if they had a subjective sense of unease, or a 'lurking doubt' about the conviction. Thus, the unsafe or unsatisfactory ground meant that the question for the Court was 'whether there is not some lurking doubt in our minds which makes us wonder whether an injustice has been done'.⁴⁹¹ The conviction in *Cooper*, based on disputed identification evidence, was subsequently quashed.

*Stafford v DPP*⁴⁹² concerned how the Court should deal with appeals raising fresh evidence. The appellants argued that the Court of Appeal should be required to decide whether the fresh evidence might have raised a reasonable doubt in the mind of the jury – the 'jury impact' test, and argued that the Court should quash the conviction if they found it did. This was said to be following Lord Parker in *Parks* (referred to above) who said that the Court was to decide whether the evidence was 'credible in the sense that it is well capable of belief; it is not for this court to decide whether it is to be believed or not'.⁴⁹³

This was rejected by the House of Lords, who held that the Court of Appeal is required to determine the impact of the fresh evidence on their own minds. This is because section 2 of the 1968 Act required the Court to quash a conviction 'if *they think*' the conviction is unsafe or unsatisfactory. Viscount Dilhorne doubted whether Lord Parker in *R v Parks* was laying down a rule of law that the Court of Appeal should consider the impact of the fresh evidence on the jury without considering what weight they give to the fresh evidence themselves.⁴⁹⁴ It was one way of determining the safety of the conviction, but the ultimate question for the Court of Appeal was whether they themselves thought the conviction was unsafe or unsatisfactory. Viscount Dilhorne doubted whether there was a great deal of difference between the 'jury impact' test and the more 'subjective test', noting: 'if

⁴⁹¹ *Cooper* (n 489) at 86.

⁴⁹² [1974] AC 878.

⁴⁹³ *R v Parks* (1962) 46 Cr App R 29

⁴⁹⁴ *Stafford* (n 492) 893.

the Court has no reasonable doubt about the verdict, it follows that the Court does not think the jury could have one'.⁴⁹⁵

The case was controversial. Lord Devlin was particularly critical of the decision, arguing that it blurred the boundary between the judge and jury.⁴⁹⁶ The issue arises primarily in unsuccessful appeals, as it meant that some people were serving periods of imprisonment despite a jury having not heard all of the evidence. Instead, judges had heard additional evidence, never seen by a jury, and decided that it could not have impacted the outcome of the trial. This was seen as an infringement of the primacy of the jury. It also appeared to be at odds with the Court's own stated reluctance not to retry appellants and reach decisions of guilt or innocence themselves. There appeared to be an uncomfortable divergence of positions – where the Court was prepared to 'usurp' the role of the jury and take it upon themselves when considering whether fresh evidence made a conviction unsafe or unsatisfactory, while at the same time stating that it was not their role. This added to the increasingly confusing issue of the proviso, the 'lurking doubt' principle, and the Court's new powers of retrial.

The 1983 Report by JUSTICE condemned the approach of the Court, summing it up as so:

'The Court has tied its own hands so that only a bad mistake by the trial judge in summing-up, some legal technicality, or fresh evidence, as narrowly defined by the 1968 Act, will result in the upsetting of a conviction'.⁴⁹⁷

It was during the operation of the 1968 Act that the 'greatest disaster to have shaken British justice'⁴⁹⁸ occurred. This 'disaster', or 'crisis',⁴⁹⁹ was a series of high-profile miscarriages of justice which had not been rectified by the Court of

⁴⁹⁵ *ibid.*

⁴⁹⁶ P Devlin, *The Judge*, (Oxford University Press 1979) 158.

⁴⁹⁷ Cited in Nobles and Schiff (n 470) 72-3.

⁴⁹⁸ Lord Devlin, 'The Conscience of the Jury' (1991) 107 (Jul) LQR 398.

⁴⁹⁹ A number of books were written on this period, also commonly called a 'crisis', see C Walker, K Starmer (eds) *Justice in Error* (Blackstone 1993); M McConville, L Bridges (eds) *Criminal Justice in Crisis* (Edward Elgar 1994).

Appeal. It culminated in numerous quashed convictions for very serious offences in the late 1980s and early 1990s.

The crisis peaked with the quashing of the convictions of the *Birmingham Six*.⁵⁰⁰ Their convictions were quashed when fresh evidence revealed that the tests carried out by Home Office scientists could not distinguish between explosive nitro-glycerine, and chemicals found in every day products, such as cigarettes, playing cards and, possibly, the soap used by the scientists to clean their porcelain bowls.⁵⁰¹ The day the Six were released from prison, the Home Secretary convened the *Royal Commission on Criminal Justice (RCCJ)*.⁵⁰² The RCCJ commissioned empirical research into the operation of the unsafe or unsatisfactory test under the 1968 Act, which was conducted by Kate Malleon.⁵⁰³ She identified the particular problem areas for the Court as being its dealing with fresh evidence,⁵⁰⁴ the lurking doubt principle⁵⁰⁵ and the proviso.⁵⁰⁶ She analysed the first 300 cases decided in 1990 and found that only four convictions were quashed on the basis of fresh evidence.⁵⁰⁷ Most grounds of appeal were procedural grounds regarding errors which occurred at trial.⁵⁰⁸ In particular there was concern that cases with fresh evidence were often rejected or treated with great caution by the Court.⁵⁰⁹ Malleon found that, regarding fresh evidence, the Court took a subjective approach to assessing it despite the criticisms levelled towards *Stafford*.⁵¹⁰ The amendments to the Court's powers in the 1960s had led to no great change in approach, which she considered to be unduly restrictive.⁵¹¹ She argued that the Court's preoccupation was in preserving finality, which could be best served by rarely reopening factual issues.⁵¹²

⁵⁰⁰ *R v McKenny & Others* [1991] 93 Cr App R 287

⁵⁰¹ *ibid*, 299-300.

⁵⁰² Royal Commission on Criminal Justice, *Report* (cm 2263, 1993) (RCCJ).

⁵⁰³ K Malleon, *The Royal Commission on Criminal Justice: Review of the Appeal Process*, Research Study No 17 (HMSO 1993).

⁵⁰⁴ *ibid*, 5.

⁵⁰⁵ *ibid*, 6.

⁵⁰⁶ *ibid*.

⁵⁰⁷ *ibid*, 9.

⁵⁰⁸ *ibid*, 8.

⁵⁰⁹ *ibid*, 11.

⁵¹⁰ *ibid*, 10.

⁵¹¹ *ibid*.

⁵¹² K Malleon, 'Appeals Against Conviction and the Principle of Finality' (1994) 21 *JL & Soc'y* 151.

The RCCJ's terms of reference were to 'examine the effectiveness of the criminal justice system in England and Wales in securing the conviction of those guilty of criminal offences and the acquittal of those who are innocent'.⁵¹³ Clearly, by its power to quash convictions, the Court has a key role to play in this. The RCCJ was concerned that the Court appeared reluctant to consider whether the jury had reached a wrong decision when an appellant could point to no procedural irregularity as having occurred at trial.⁵¹⁴ This was based partly upon Malleson's empirical research which showed that appeals based on fresh evidence or lurking doubt (i.e. appeals not raising procedural irregularities) were unlikely to be successful when compared to procedural irregularity appeals.

The Commission's recommendations led to the enactment of the 'unsafety test', via amendment of the 1968 Act by the *Criminal Appeal Act 1995*. The 1995 Act also created the *Criminal Cases Review Commission* (CCRC), as recommended by the Commission. The CCRC has the power to refer appeals back to the Court of Appeal if it thinks there is a 'real possibility' that the Court will find the conviction to be unsafe.⁵¹⁵ The Court's power to receive fresh evidence is now contained in section 23 of the 1968 Act (as amended), which states that the Court has the power to receive fresh evidence if it thinks it is in the interests of justice to do so. When deciding whether it is in the interests of justice the Court should consider, (section 23(2)(a)), whether the evidence appears to the Court to be capable of belief; (b) whether it appears to the Court that the evidence may afford any ground for allowing the appeal; (c) whether the evidence would have been admissible in the proceedings from which the appeal lies; and (d) whether there is a reasonable explanation for the failure to adduce the evidence in those proceedings. Finally, the 'proviso', which allowed the Court to dismiss an appeal if it thought no miscarriage of justice had occurred, was abolished.

⁵¹³ RCCJ (n 502) i.

⁵¹⁴ *ibid*, Chapter 10 [3].

⁵¹⁵ *Criminal Appeal Act 1995*, s 13.

4.4 Analysis of the ‘unsafety test’

Dennis⁵¹⁶ argued that the fundamental function of the Court since the enactment of the ‘unsafety test’ is to review the legitimacy of convictions.⁵¹⁷ In particular, this means that the operation of the test is based upon whether the particular judges hearing the appeal, at the particular time that the appeal is heard, are sure of the factual accuracy of the conviction; its moral authority; and that the conviction is grounded in the rule of law.⁵¹⁸ It is only if the Court can answer all of these questions affirmatively that convictions can be safe.

As can be seen from the history of the Court discussed in the previous section, it has been by design that the Court’s powers have become progressively less explicit and somewhat more vague or open-textured. Under the 1907 Act, the Court was constrained to only allow appeals in the absence of procedural irregularities if there was no evidence upon which a jury could have convicted. Since the jury did convict, it may be understood why this was hardly ever applicable. If the jury convicted when there was insufficient evidence, this may constitute an error of the trial judge for not stopping the case on the basis of no case to answer.

The RCCJ called upon the Court to generally be more ready to reverse jury verdicts than had been previously, and the ‘unsafety test’ was the way chosen to permit the Court the powers to do so whenever it thinks it just.⁵¹⁹ Thus, whilst the ‘unsafety test’ was designed to give the Court more general powers, it was done so with a proviso, or a hope / expectation, that it would exercise that discretion in a particular way; namely, in a more liberal way. As will be explained below, this hope appears to have failed, as the test itself does not constrain the judges to deciding cases in a liberal way. The test may be considered so ‘open-textured’ that it is has been necessary for the Court to interpret what it means, and to impose certain rules within its operation. The extent to which the Court then

⁵¹⁶ I Dennis, ‘Fair Trials and Safe Convictions’ [2003] CLP 211

⁵¹⁷ *ibid*, 236.

⁵¹⁸ *ibid*.

⁵¹⁹ RCCJ (n 502) Chapter 10 [1].

follows those rules in the operation of the test, is how the law governing appeals has been measured in this study.

There have been two empirical studies of the Court since the enactment of the 'unsafety test'. Roberts's research was structured in 'the qualitative method in conjunction with a purely descriptive quantitative analysis'.⁵²⁰ She presented a number of factors which could indicate the approach of the Court. These were: the success rate of appeals; the numbers of applications for leave which were granted; how the Court dealt with fresh evidence and lurking doubt appeals; its approach to procedural irregularity appeals, including issues under the *Human Rights Act 1998*; and the use of its powers to order a retrial.⁵²¹ She found that a falling overall success rate, and the low number of successful appeals on fresh evidence and lurking doubt appeals to be some evidence of a restrictive approach.⁵²² She found that despite the RCCJ's view that the Court of Appeal should be more open to quashing convictions in the absence of procedural irregularities, such appeals were rarely successful.⁵²³ Across her sample, only one appeal was allowed on the basis of the Court finding a 'lurking doubt', and only nine were allowed on the basis of fresh evidence.⁵²⁴ She concluded that attempts to liberalise the Court's practice had failed, owing to the Court's function of reviewing convictions rather than retrying appellants.⁵²⁵

Heaton's study⁵²⁶ reached similar conclusions. He also found that fresh evidence and lurking doubt appeals were rarely successful.⁵²⁷ He found that the Court appeared to use the fresh evidence provisions in a restrictive way, 'thus limiting

⁵²⁰ *ibid.*

⁵²¹ S Roberts 'The Decision-Making Process of Appeals Against Conviction in the Court of Appeal (Criminal Division) (DPhil Thesis, London School of Economics and Political Science, 2009) 209-11. (Hereafter 'Roberts (2009)').

⁵²² *ibid.*

⁵²³ *ibid.*

⁵²⁴ *ibid.*

⁵²⁵ S Roberts 'The Royal Commission on Criminal Justice and Factual Innocence: Remedying Wrongful Convictions in the Court of Appeal' (2004) (1)(2) JJ 86. (Hereafter 'Roberts (2004)').

⁵²⁶ SJ Heaton, *A Critical Evaluation of Using Innocence as a Criterion in the Post-Conviction Process* (DPhil Thesis, University of East Anglia, 2013).

⁵²⁷ *ibid.*, 125.

the number of successful appeals'.⁵²⁸ He further contended that the Court of Appeal 'explicitly eschews interest in innocence in a significant number of cases. It also, by its reluctance to adopt a less restrictive approach to the receipt and evaluation of fresh evidence, represents a significant obstacle to those asserting innocence'.⁵²⁹

These studies of the Court of Appeal's decision-making has evaluated the Court from a particular perspective – what may be called the 'wrongful conviction' or 'miscarriage of justice' perspective.⁵³⁰ Naughton argues that writers from the miscarriage of justice community can 'identify the key difficulties ... [for] the delivery of justice for innocent victims of wrongful convictions' and form a 'counter-discourse on the existing arrangements'.⁵³¹ In relation to the Court of Appeal, the difficulty is that it is said to have continued to have adopted a restrictive approach following the enactment of the 'unsafety test'. Although the Court has been criticised for not adapting its approach following the adoption of the 'unsafety test', it was suggested above that it should not have been surprising that the test does not appear to have liberalised the Court's approach.

This thesis approaches the analysis of the Court of Appeal from a differing perspective to previous studies. This study is embedded in the ELS community. It may be said that this research complements, and is complemented by, previous research, but they ask different questions. Previous studies have sought to discern the Court's approach to determining appeals; this thesis asks whether the Court has the legitimacy to render decisions at all by holding the Court to a standard of impartiality. Moreover, whilst previous research suggests that a 'restrictive approach' is what leads some appeals to be allowed and some to be dismissed, this thesis asks whether there is an association between a range of

⁵²⁸ *ibid*, 117.

⁵²⁹ *ibid*, 211.

⁵³⁰ M Naughton (ed), *The Criminal Cases Review Commission: Hope for the Innocent?* (Palgrave Macmillan 2009) 13.

⁵³¹ *ibid*.

variables and outcomes. It is discussed in Chapter 8 whether the concept of an 'approach' is a solid enough foundation for empirical research.

This study is not concerned with the same aspects of the Court's decision-making as are some previous studies. This thesis is not concerned with analysing decisions of particular cases and determining whether the outcomes of appeals are right or wrong. Moreover, it does not address the question of how well the Court performs in correcting miscarriages of justice, or what its approach to correcting miscarriages of justice is. What is involved in this study is a 'broadening of the traditional analytical approach'.⁵³² This entails foregoing the ability to judge the nuances of particular cases in favour of a broad perspective on fact patterns and decision-making.⁵³³ It is not necessary to provide a definition of a miscarriage of justice, because whether a miscarriage of justice occurred or was rectified (or not) in a particular appeal is irrelevant to the analysis conducted here. As such, there is no definition of a miscarriage of justice offered, and this thesis is neutral as to how well the Court performs in correcting miscarriages of justice, and has not sought to directly discern evidence of the Court's approach.

The previous research on the Court of Appeal is useful in explaining the institutional position of the Court. Studies which analyse courts from the perspective of the 'institutional model' are interested in the extent to which the institutional norms of a court mediate judges' preferences, and also how the law guides decision-making.⁵³⁴ In the Court of Appeal, the law which guides decision-making is the interpretation of 'unsafety test'. There has been a considerable amount of jurisprudence from the Court of Appeal as to the meaning of unsafety and its relationship with other concepts, such as unfairness. It is these norms which are captured within the legal variable which is analysed in the binary logistic

⁵³² A Juliano and SJ Schwab 'The Sweep of Sexual Harassment Cases' (2001) (86) Cornell L Rev 548, 553.

⁵³³ *ibid.*

⁵³⁴ See K Weinshall-Margel, 'Attitudinal and Neo-Institutional Models of Supreme Court Decision Making: An Empirical and Comparative Perspective From Israel' (2011) 8 JELS 556, 558.

regression models shown in Chapter 7. The remainder of this chapter analyses the ‘unsafety test’.

4.5 The meaning of unsafety

As will be explained below, some kinds of appeals are automatically unsafe.⁵³⁵ For the rest, and this is the majority, the unsafety test is an example of counterfactual reasoning.⁵³⁶ This means that the Court must take what did happen (all the circumstances of the trial and conviction), and decide what would have happened if some other hypothetical events had occurred. For instance, if an error had not occurred, or if the ‘fresh evidence’ had been available at trial. The Court must determine whether what went wrong at trial leads them to believe that the conviction is unsafe, and the usual way to do this is to ask whether they think a guilty verdict would still have been returned if the error had not occurred or the jury had heard the fresh evidence. Thus, the Court frequently will operate a ‘jury impact’ test, despite the Court frequently reiterating that its test is the ‘unsafety test’. The ‘unsafety test’ could be understood as a hypothesis test similar in character to those employed in this thesis. In appeals against conviction, the null hypothesis is that the conviction is safe. The conviction will only be unsafe if there is sufficient doubt that the conviction is just or legitimate.

The word ‘unsafe’ appears to carry little independent meaning itself. Cohen famously referred to the ‘thingification’ of legal concepts.⁵³⁷ By this he meant that legal concepts, such as *property*, do not create rights themselves but ‘merely [recognise] a pre-existent Something’.⁵³⁸ As the review of the history of the Court showed, the ‘unsafety test’, as well as the previous tests, was introduced to seek to assist the Court in achieving the inchoate ends of doing justice. This appears to ‘thingify’ the concept of doing justice in individual cases. The ‘unsafety test’

⁵³⁵ In particular, ‘*Re Bennett*-like’ abuses of process are automatically unsafe. See H Blaxland, ‘Sappers and Underminers: Fresh Evidence Revisited’ (2017) Crim LR 537, 542.

⁵³⁶ See DA Winkelmann et. al. ‘An Empirical Method for Harmless Error’ (2014) (46) Ariz State L J 1405.

⁵³⁷ FS Cohen, ‘Transcendental Nonsense and the Functional Approach’ (1935) Colum L Rev 809, 815.

⁵³⁸ *ibid.* (As in original).

appears open-textured, but the discretion of judges is not wholly unfettered. If the unsafety test did offer unfettered discretion, judges would be able to decide cases however they wished, and the law would play little role. This would be indicative of, as attitudinal researchers argued, the law being a 'cloak' for judges pursuing policy goals. However, as discussed below, the discretion is tempered by rules.

The jurisprudence of the Court provides guidance, based upon legal rules, as to how the test should be exercised in certain circumstances. Mantell LJ in *R v Davis, Rowe, and Johnson*⁵³⁹ (the case of the *M25 Three*), said that the Court should apply the principle from *Stirland v DPP*⁵⁴⁰ when determining whether a conviction is unsafe. *Stirland* was authority for how the 'proviso' should be utilised under the 1907 and 1968 Acts. The House of Lords in *Stirland* applied the proviso and upheld the conviction because 'there was an overwhelming case proved against the appellant ... no reasonable jury, after a proper summing up, could have failed to convict the appellant'.⁵⁴¹ In *Woolmington v DPP*⁵⁴² the House of Lords declined to apply the proviso and the conviction was quashed, because 'we cannot say that if the jury had been properly directed they would have inevitably come to the same conclusion'.⁵⁴³

To apply the proviso the Court had to be sure that 'no reasonable jury could have failed to convict' or be sure that the jury would 'inevitably come to the same conclusion'. Mantell LJ in *Davis* adopted this test and stated that the Court had to consider 'would a reasonable jury have been bound to return verdicts of guilty?'.⁵⁴⁴ Under the proviso, if the only reasonable verdict was one of guilty the proviso would be applied and the conviction upheld; under the 'unsafety test' the conviction will simply be safe. Thus, whilst the test for the Court is always whether

⁵³⁹ [2001] 1 Cr App R 115.

⁵⁴⁰ [1944] AC 315.

⁵⁴¹ *ibid*, at 321.

⁵⁴² [1935] AC 462.

⁵⁴³ *ibid* at 482-3.

⁵⁴⁴ *Davis* (n 539) at 145.

the conviction is unsafe, the Court often in effect must place itself in the position of the jury and try to predict what the outcome would have been.

An early case decided under the 'unsafety test' is *R v CCRC ex parte Pearson*.⁵⁴⁵ Lord Bingham acknowledged that 'trials by judge and jury may on occasion result in wrongful convictions'⁵⁴⁶ and that the Court of Appeal 'exists to correct such errors in appeals brought before it'.⁵⁴⁷ The expression 'unsafe,' he said, 'does not lend itself to precise definition' but:

'In some cases unsafety will be obvious, as (for example) where it appears that someone other than the appellant committed the crime ... or where a conviction is shown to be vitiated by serious unfairness in the conduct of the trial or significant legal misdirection ... Cases however arise in which unsafety is much less obvious: cases in which the Court, although by no means persuaded of an appellant's innocence, is subject to some lurking doubt or uneasiness whether an injustice has been done. If ... the Court entertains real doubts whether the appellant was guilty of the offence, the Court will consider the conviction unsafe'.⁵⁴⁸

From this passage, it will be seen that if the Court thinks somebody else committed the offence, i.e. 'is innocent', or even merely entertains 'real doubts' or a 'lurking doubt' about guilt, the conviction will be unsafe. This shows that the Court clearly does have an interest in determining whether an appellant is innocent. Naughton is critical of this passage and suggests that Lord Bingham's statement is a 'highly misleading form of judicial communication to the public'.⁵⁴⁹ He utilises the cases of *Stefan Kiszko* and *Sean Hodgson* as evidence of the Court taking a legalistic approach rather than an approach focussed upon correcting the conviction of the factually innocent.⁵⁵⁰ Naughton states that these convictions were not overturned because they were factually innocent, but

⁵⁴⁵ (2000) 1 Cr App R 141.

⁵⁴⁶ *ibid*, 146.

⁵⁴⁷ *ibid*.

⁵⁴⁸ *ibid*, 146-7.

⁵⁴⁹ M Naughton, *The Innocent and the Criminal Justice System: A Sociological Analysis of Miscarriages of Justice* (Palgrave Macmillan 2013) 152.

⁵⁵⁰ *ibid*, 154.

because they were able to show a breach of process or produce fresh evidence that was not available at the time.

This argument is not persuasive as Naughton does not appear to appreciate that the convictions were quashed because the fresh evidence *proved* (as far as possible) the appellants were innocent at the time of the appeal. In Lord Judge in *Hodgson*⁵⁵¹ stated that there was no police misconduct, no untruthful or mistaken witnesses, and nothing done by anybody at trial could be criticised.⁵⁵² The conviction was quashed because fresh DNA evidence destroyed any possible link between Hodgson and the murder victim, leading to the conclusion that somebody else must have been the killer.⁵⁵³ This is in contrast with *Davis*, where at the time of the appeal their innocence was not proved, and so the conviction had to be quashed only on the basis of the police malpractice.

It will be observed that Lord Bingham's understanding of the unsafety test means that convictions can be unsafe for a wide variety of reasons, including belief in innocence, unfair trials, or a lurking doubt. All these grounds for finding a conviction unsafe adhere to Dennis's argument that the judges must be satisfied in the legitimacy of convictions if they are to be safe. With this general understanding of the 'unsafety test' in mind, the next section explains under what circumstances convictions will be unsafe on the basis of unfairness or procedural irregularities.

4.6 Procedural irregularities and unfairness

Although the concept of a fair trial far predates 1998, the *Human Rights Act 1998* incorporated the *European Convention on Human Rights* into English law. The effect of the Human Rights Act 1998 is that a breach of Article 6 can be argued in British courts, for instance, as a ground for quashing a conviction. This raises the question of the relationship between unsafety and unfairness. It is

⁵⁵¹ [2009] EWCA Crim 490.

⁵⁵² *ibid*, [5].

⁵⁵³ *ibid*, [45].

uncontroversial that convictions can be unsafe even if a trial was fair.⁵⁵⁴ This can arise when the Court has any lurking, or greater, doubt about the factual accuracy of the verdict (or is sure that it is factually inaccurate). As stated in *Hodgson*, discussed above, under such circumstances the conviction will be unsafe, even if the trial was entirely properly conducted and fair.

The question of whether a finding that a trial was unfair trial will always render a conviction unsafe is far more complex. As Dennis explained, the Court initially answered this question in four different ways.⁵⁵⁵ The case of *R v Forbes*⁵⁵⁶ adopted an 'absolutist'⁵⁵⁷ position, in that it was stated that if 'a defendant's right to a fair trial has been infringed, a conviction *will* be held to be unsafe'.⁵⁵⁸ A 'quasi-absolutist'⁵⁵⁹ position was adopted by the Court of Appeal in *R v Togher*.⁵⁶⁰ In *Togher* the Court of Appeal said 'if a defendant has been denied a fair trial it will be *almost inevitable* that the conviction will be regarded as unsafe'.⁵⁶¹ 'Almost inevitable' represents a retreat from the absolutist position, but suggests that an unfair trial will be likely render a conviction unsafe. In *Togher*, the reason that the unfairness did not make the conviction inevitably unsafe appears to be that the appellants had pleaded guilty to the offence, and the unfairness did not influence their decision to plead guilty. There was thus no reason to doubt the factual accuracy of the verdict.

Some cases adopted a more cautious approach, in which whether an unfair trial made a conviction unsafe was contingent upon all the circumstances of the case.⁵⁶² An example of this 'contingent position' is *R v Davis, Rowe and Johnson*.⁵⁶³ In that case, the appellants had received a ruling from the *European*

⁵⁵⁴ See I Dennis, 'Fair Trials and Safe Convictions' [2003] CLP 211, 218.

⁵⁵⁵ *ibid*, 219.

⁵⁵⁶ [2001] 1 AC 473.

⁵⁵⁷ Dennis (n 554) 212.

⁵⁵⁸ *Forbes* (n 556) at [24]. (Emphasis added).

⁵⁵⁹ Dennis (n 554) 212.

⁵⁶⁰ *R v Togher; R v Doran; R v Parsons* [2001] 1 Cr App R 33

⁵⁶¹ *ibid*, [30]. Emphasis added.

⁵⁶² Dennis (n 554) 212-3.

⁵⁶³ [2001] 1 Cr App R 8.

Court of Human Rights that the trial had been unfair.⁵⁶⁴ In the Court of Appeal, Mantell LJ said it was the nature and degree of the unfairness which determined whether the conviction was unsafe and this would depend upon the circumstances of the particular case.⁵⁶⁵ He explicitly '[rejected] ... the contention that a finding of a breach of Article 6(1) by the ECHR leads inexorably to the quashing of the conviction'.⁵⁶⁶

The fourth and final position identified by Dennis is that there is no relationship between unsafety and unfairness unless the unfairness leads to doubt about the factual accuracy of the verdict.⁵⁶⁷ This is derived from Auld LJ's judgment in *R v Chalkley*,⁵⁶⁸ which was decided prior to the enactment of the *Human Rights Act 1998*. The appellants had pleaded guilty based on recordings obtained when covert listening devices were illicitly placed in their homes. The trial judge accepted that the instillation of the listening devices was illegal.⁵⁶⁹ Auld LJ said that, despite the 'unsafety test' being designed to induce the Court to be more liberal, 'the new provision ... may be ... narrower than before'.⁵⁷⁰ He held that the deletion of the 'unsatisfactory' part of the test meant that the Court now 'has no power to allow an appeal if it does not think the conviction unsafe but is dissatisfied in some way with what went on at the trial'.⁵⁷¹ Accordingly, the appeal was dismissed because 'by their guilty pleas, they intended to admit their guilt, and that their convictions are, therefore, safe'.⁵⁷²

The decision in *Chalkley* could be considered a narrow interpretation of the test. It meant that the Court would have very little power to oversee the conduct of State officials. This is despite, in *R v Horseferry Road Magistrates' Court ex parte*

⁵⁶⁴ (2000) 30 EHRR 1.

⁵⁶⁵ *Davis* [2001] 1 Cr App R 115 at [65].

⁵⁶⁶ *ibid.*

⁵⁶⁷ Dennis (n 554) 219-20.

⁵⁶⁸ [1998] 2 Cr App R 79.

⁵⁶⁹ *ibid.*, 87.

⁵⁷⁰ *ibid.*, 98.

⁵⁷¹ *ibid.*

⁵⁷² *ibid.*, 100.

Bennett,⁵⁷³ Lord Griffiths accepting on behalf of the judiciary the ‘responsibility for the maintenance of the rule of law that embraces a willingness to oversee executive action’.⁵⁷⁴ The decision in *Chalkley* must now be considered to have been decided incorrectly on principle. The Court of Appeal in *R v Mullen*⁵⁷⁵ held that a conviction can be unsafe even if the Court is sure of its factual accuracy, because ‘for a conviction to be safe, it must be lawful’.⁵⁷⁶ Mullen was convicted of conspiracy to cause explosions due to being part of an IRA cell, and was sentenced to 30 years’ imprisonment. After serving ten years imprisonment, he discovered that he had been illegally extradited from Zimbabwe (to where he had fled) following collusion between the British and Zimbabwean secret services. The Court found this to be a ‘blatant and extremely serious failure to adhere to the rule of law’ and so quashed the conviction’.⁵⁷⁷

Dennis argued that the relationship between unfairness and unsafety is that if the unfairness is so severe that the conviction loses its legitimacy, such as *Mullen*, the conviction will be quashed.⁵⁷⁸ Most appeals do not concern irregularities as serious as this, and so are concerned with the effect of a violation on the question of the safety of the conviction.⁵⁷⁹ In particular, the effect must be that the outcome of the trial could reasonably have been different.⁵⁸⁰ As was stated above, this involves an element of counterfactual reasoning as to what would have happened if the irregularity had not occurred. Thus, whilst the ‘unsafety test’ provided the Court with a level of discretion to determine its own powers, it has indicated the circumstances in which unfairness or irregularities will render a conviction unsafe. This, it is argued, is now part of the unsafety test.

Some procedural irregularities, if found to have occurred, as a matter of logic must render convictions unsafe. If the error is of a type that the trial should have

⁵⁷³ [1994] 1 AC 42.

⁵⁷⁴ *ibid*, 62.

⁵⁷⁵ [1999] 2 Cr App R 143.

⁵⁷⁶ *ibid*, 161.

⁵⁷⁷ *ibid*, 156-7.

⁵⁷⁸ *Dennis* (n 554) 214.

⁵⁷⁹ *ibid*.

⁵⁸⁰ *ibid*.

been stopped before a verdict was reached, it follows that any verdict must be unsafe. This applies when, for instance, the Court decides that the trial judge should have accepted an application of no case to answer, or should have accepted an application to stay the proceedings. In *R v Smith*⁵⁸¹ Mantell LJ opined that if the Court decides that the trial judge should have accepted an application of no case to answer made after the completion of the prosecution case, the conviction must be quashed, even if something is said later which proves the appellant's guilt.⁵⁸² In *R v Broadhead*⁵⁸³ a conviction for murder was quashed because 'it follows that, but for the ruling, the jury's verdict would have been different. It would have been to acquit the defendant on the direction of the judge'.⁵⁸⁴

For procedural irregularity errors which are not presumptively unsafe, or do not make the trial unfair, the Court exercises a higher degree of discretion under the unsafety test. It must also operate the 'jury impact' test. If a procedural irregularity is found to have occurred, the conviction is only *prima facie* unsafe. A further step is required; it must be decided whether 'the outcome of the trial might have been different but for the irregularity'.⁵⁸⁵ The same approach is observed by Spencer, who notes that usually the Court 'will uphold the conviction if it is convinced that the defendant is really guilty, and would still have been convicted even if the irregularity had not taken place'.⁵⁸⁶

The Court has frequently reiterated that this is a two-step process: if an error was found the Court must still usually decide whether the error makes the conviction unsafe. This two stage test can be seen in *R v Beedall* where the appeal against a conviction for rape was dismissed as 'no injury was done' to the safety of the conviction by the trial judge's error in summing up.⁵⁸⁷ In *R v Jheeta*, where the

⁵⁸¹ [1999] 2 Cr App R 238.

⁵⁸² *ibid*, 242. See also *R v Fletcher* [2009] EWCA Crim 1187 at [3].

⁵⁸³ [2006] EWCA Crim 1705.

⁵⁸⁴ *ibid*, [14].

⁵⁸⁵ Dennis (n 554), 213.

⁵⁸⁶ JR Spencer, 'Quashing Convictions for Procedural Irregularities' (2007) Crim LR 835, 837.

⁵⁸⁷ [2007] EWCA Crim 23, [28].

Court found that the appellant was wrongly advised by his legal advisors to plead guilty to rape, the conviction was safe as the Court ‘entertained no reservations’ that he was, in fact, guilty of rape and his guilty plea reflected that fact.⁵⁸⁸ In *R v Dada*⁵⁸⁹ the appeal against convictions for rape was dismissed despite there being a ‘number of unsatisfactory features about [the] case’ as ‘none of these matters bear on the safety of the conviction, they are concerned with the proper conduct of the trial.’⁵⁹⁰ The relationship between the finding of an error and the outcome of an appeal against conviction is an important consideration in this thesis. This is because the presence of an error, and its effect on the safety of a conviction, has been utilised as a measurement of the law orbiting appeals.

4.7 Fresh evidence appeals

Fresh evidence appeals raise different issues to appeals based upon procedural irregularities or an unfair trial. This is because they are based upon factual evidence which a jury has never seen. In *R v Pendleton*⁵⁹¹ the House of Lords affirmed the subjective *Stafford* approach was the correct way to assess fresh evidence.⁵⁹² In *Pendleton* appellant argued that ‘it is not permissible for appellate judges ... to make their own decision on the significance or credibility of the evidence’.⁵⁹³ He had argued that the Court should always allow an appeal if it thought the fresh evidence *might* raise a reasonable doubt in the mind of the jury. This ‘jury impact’ test might be considered a more liberal test because it may prevent the judges dismissing appeals if they personally remain sure of guilt when credible fresh evidence is produced. It may also be said that the jury impact test preserves the normative position of the jury, by requiring convictions to be quashed if credible fresh evidence, never seen by a jury, is uncovered.

⁵⁸⁸ [2007] EWCA Crim 1699 [28-9].

⁵⁸⁹ [2008] EWCA Crim 2121 [21-4]. There are numerous other examples available, for example *R v Wood* [2008] EWCA Crim 587; *R v Gibson* [2006] EWCA Crim 542; *R v Ramirez* [2009] EWCA Crim 1721.

⁵⁹⁰ *Dada*, *ibid*, [21-4].

⁵⁹¹ [2001] UKHL 66.

⁵⁹² *ibid*, [19].

⁵⁹³ *ibid*, [12].

Lord Bingham rejected that the 'jury impact' test should be used to determine fresh evidence appeals. He pointed out that it is 'anomalous for the Court to quash a conviction if it raised no doubt whatever in their minds but might have raised a reasonable doubt in the minds of the jury'.⁵⁹⁴ However, he suggested that recourse to the 'jury impact' test might be appropriate in certain circumstances:

'it will usually be wise for the Court of Appeal, *in a case of any difficulty*, to test their own provisional view by asking whether the evidence, if given at the trial, might reasonably have affected the decision of the trial jury to convict. If it might, the conviction must be thought to be unsafe'.⁵⁹⁵

This could be considered a narrower interpretation of the fresh evidence provisions, because it could be argued that Lord Bingham sought to reduce the scope of the 'jury impact' test by limiting its use to 'cases of difficulty'.⁵⁹⁶ This criticism is unwarranted because it is difficult to foresee many contested cases in the Court of Appeal which would not be a 'case of difficulty'.⁵⁹⁷ Lord Bingham allowed the appeal and quashed the conviction because, although the Court of Appeal had applied the correct test:

'In the light of ... this fresh psychological evidence it is impossible to be sure that this conviction is safe, and *that is so whether the members of the House ask whether they themselves have reason to doubt the safety of the conviction or whether they ask whether the jury might have reached a different conclusion* ... In holding otherwise the Court of Appeal strayed beyond its true function of review and made findings which were not open to it in all the circumstances. Indeed, it came perilously close to considering whether the appellant, in its judgment, was guilty'.⁵⁹⁸

⁵⁹⁴ *ibid.*

⁵⁹⁵ *ibid.* (Emphasis added).

⁵⁹⁶ See SJ Heaton, *A Critical Evaluation of Using Innocence as a Criterion in the Post-Conviction Process* (DPhil Thesis, University of East Anglia, 2013) 194.

⁵⁹⁷ See H Blaxland and P Wilcock 'Fresh Evidence in Criminal Appeals – Pendleton Revisited' (2006) *Arch News* 4.

⁵⁹⁸ *Pendleton* (n 591) [28] (Emphasis added).

It appears that Lord Bingham was conferring a power upon the Court to quash convictions on the basis of fresh evidence if it is not satisfied that the jury would still have convicted. It can reach this conclusion *either* because the fresh evidence causes the Court to think the conviction is unsafe, *or* because it thinks it may have impacted the jury. What Lord Bingham was emphasising in *Pendleton* is that, however the decision is reached, the primary test for the Court is unsafety. The confusion which arises is that whilst judges avoid admitting to utilising the jury impact test in fresh evidence appeals, the jury impact test, as discussed above, is integral to the unsafety test itself. This is concomitant with the counterfactual nature of the ‘unsafety test’ – the judges must decide what would have happened if some counterfactual state had occurred. In the case of fresh evidence, this will often mean the judges must decide what would have happened if the jury had been aware of the evidence in order to decide whether it is unsafe.

Lord Hobhouse in *Pendleton* said that ‘in my judgment it is not right to attempt to look into the minds of the members of the jury ... it is for the Court of Appeal to answer ... do we think that the conviction was unsafe?’⁵⁹⁹ He agreed that the conviction should have been quashed, but this was not due to the fresh evidence but because the verdict was inconsistent with the directions of the trial judge.⁶⁰⁰ While Lord Bingham left open the ‘jury impact’ test (at the same time clarifying that it was not the sole test), Lord Hobhouse rejected the appellant’s position that he was ‘seeking to escape from the verdict of a jury merely upon the possibility (which will exist in almost every case) that the jury might have returned a different verdict’.⁶⁰¹

Lord Hobhouse’s view appeared to have been followed in a number of cases. In *Dial and Dottin v Trinidad and Tobago*,⁶⁰² Lord Brown stated that: ‘the primary question is for the Court itself and is not what effect the fresh evidence would

⁵⁹⁹ *ibid*, [38]

⁶⁰⁰ *ibid*, [40].

⁶⁰¹ *ibid*, [36].

⁶⁰² [2004] UKPC 4.

have had on the mind of the jury'.⁶⁰³ He added that 'the question arising for the Appeal Court's determination is whether [the fresh evidence] realistically places the appellant's guilt in reasonable doubt'.⁶⁰⁴ It could be argued that this is narrower, because the Court stressed the subjective approach and may have relegated the 'jury impact' test to an 'optional safeguard'.⁶⁰⁵ As Blaxland and Wilcock argued, this statement appears to contradict Lord Bingham's statement in *Pendleton* that the Court's role is to assess safety, not guilt.⁶⁰⁶ It is submitted, however, that *Dial* is consistent with *Pendleton* and that both adopt the jury impact test. As Lord Brown in *Dial* stated: 'if the Court regards the case as a difficult one, it may find it helpful to test its view "by asking whether the evidence, if given at the trial, might reasonably have affected the decision of the trial jury to convict"'.⁶⁰⁷ This clearly does not rule out the 'jury impact' test.

Dial was a majority decision, with Lord Steyn in the minority, stressing that the question for the Court was what the jury would have made of the fresh evidence.⁶⁰⁸ The dispute between the majority and the minority in *Dial* was not the nature of the test, which both endorsed Lord Bingham's statement in *Pendleton*, but what factors will make a case a 'case of difficulty'. It is submitted that majority found that it was not a case of difficulty because the fresh evidence did not realistically place the appellants' guilt in any reasonable doubt (i.e. they were sure the jury would still have convicted), while the minority thought it was a case of difficulty and the fresh evidence could have impacted the decision of the jury.

As Ashworth and Redmayne say, while it may appear that 'all is chaos',⁶⁰⁹ the law is clear. It can appear to be chaos, but that is because the outcomes of

⁶⁰³ *ibid*, [31].

⁶⁰⁴ *ibid*, [42].

⁶⁰⁵ H Blaxland and P Wilcock 'Fresh Evidence in Criminal Appeals – Pendleton Revisited' (2006) Arch News 4.

⁶⁰⁶ *ibid*.

⁶⁰⁷ *Pendleton* (n 591) [31].

⁶⁰⁸ *Dial* (n 602) [52].

⁶⁰⁹ A Ashworth and M Redmayne, *The Criminal Process* (4th ed, Oxford University Press 2010) 383.

appeals depend upon the opinions of the judges as to the overall strength of the case.⁶¹⁰ As they state, if the Court is sure of the safety or unsafety of the conviction, it will not need to apply the 'jury impact' test.⁶¹¹ It is submitted that this is the correct approach for the Court to take, if fulfilling its role in reviewing whether it is satisfied that the conviction is factually accurate, and this is what it does do.

Recently, the Court of Appeal in *R v Garland*⁶¹² has reaffirmed *Pendleton*. Namely, Lloyd Jones LJ stated that 'the ultimate question for our consideration is whether the material causes us to doubt the safety of the conviction'.⁶¹³ However, in addressing the question of unsafety, they had 'regard to the question of what impact the withheld material might have had on the jury'.⁶¹⁴ Whilst this may appear to be further reducing the scope of the 'jury impact' test, as Blaxland noted most cases in which fresh evidence is received must be considered 'cases of difficulty' at least requiring consideration of what impact the fresh evidence would have had on the jury.⁶¹⁵

The controversy regarding the Court's reception of fresh evidence is perhaps best explained by Hughes LJ's comments in *R v Ahmed*.⁶¹⁶ In this case Hughes LJ explained clearly why the 'jury impact' test cannot be the determinative test in fresh evidence appeals, but should instead be a confirmatory test of the judges' views. If the 'jury impact' test means that the Court should quash the conviction if the fresh evidence might have influenced the jury, then it is likely that all fresh evidence appeals would be successful. This is because 'it will be impossible to be 100% sure that [the fresh evidence] might not have had *some* impact on the jury's deliberations, since, *ex hypothesi* the jury has not seen the fresh material'.⁶¹⁷ Thus, as Blaxland states, the correct test must be whether the jury

⁶¹⁰ *ibid.*

⁶¹¹ *ibid.*

⁶¹² [2016] EWCA Crim 1743.

⁶¹³ *ibid* [55].

⁶¹⁴ *ibid.*

⁶¹⁵ H Blaxland, 'Sappers and Underminers: Fresh Evidence Revisited' (2017) Crim LR 537, 542.

⁶¹⁶ [2010] EWCA Crim 2899.

⁶¹⁷ *ibid* [24]. (Emphasis original).

might reasonably have not convicted if they had received the fresh evidence.⁶¹⁸ This is, as seen, the ‘unsafety test’ as explained by Mantell LJ in *Davis*, and discussed above.

Before deciding the effect of any fresh evidence the Court must decide whether to formally receive it. This is quite an artificial process because the judges will usually hear the evidence *de bene esse* before formally deciding to admit it. The Court has the power to receive fresh evidence if it thinks it is ‘necessary in the interests of justice’, by section 23 of the *Criminal Appeal Act 1968*. In determining whether it is in the interests of justice, the Court is required to consider: 23(2)(a) whether the evidence is capable of belief; (b) whether the evidence may afford a ground for allowing the appeal; (c) whether the evidence would be admissible; and (d) whether there is a reasonable explanation for failing to adduce the evidence at trial. These tests are not determinative of whether it is in the interests of justice to receive the evidence but are designed to assist the Court in deciding whether it is in the interests of justice.⁶¹⁹ As was made clear in *R v Erskine*, if the Court thinks that the fresh evidence makes the conviction unsafe it will always be in the interests of justice to receive it, even if it was technically not ‘fresh’.⁶²⁰

4.8 Lurking doubt appeals

When the concept of quashing convictions on the basis of a ‘lurking doubt’ was created by Lord Widgery in *R v Cooper*,⁶²¹ it was envisaged that the Court would consider whether there was some ‘subjective sense of unease ... which makes us wonder whether an injustice has been done’.⁶²² Recently, the Court of Appeal in *R v Pope*⁶²³ expanded upon the doctrine. Lord Judge said that:

⁶¹⁸ Blaxland (n 615) at 542.

⁶¹⁹ *R v Erskine; R v Williams* [2009] EWCA Crim 1425 at [39].

⁶²⁰ *ibid.*

⁶²¹ (1969) 53 Cr App R 82.

⁶²² *ibid.*, 86.

⁶²³ [2012] EWCA Crim 2241.

'It is not open to the Court to set aside the verdict on the basis of some collective, subjective judicial hunch that the conviction is or may be unsafe ... "lurking doubt" requires reasoned analysis of the evidence or the trial process, or both, which leads to the inexorable conclusion that the conviction is unsafe'.⁶²⁴

This was based on Leigh's article which suggested that the phrase 'lurking doubt' was actually little more than a rhetorical flourish on the unsafety test.⁶²⁵ Leigh argued that in cases purporting to have been successful on the basis of an 'inchoate hunch' (a 'lurking doubt') were in fact explainable on other grounds.⁶²⁶ Leigh argued that it was 'more principled' for the Court to take a narrower approach which respects the position of the jury.⁶²⁷

As discussed above, research on the decision-making of the Court has found that appeals based on 'lurking doubt' are unlikely to be successful. If the 'unsafety test' is understood as requiring counterfactual reasoning it becomes clearer why this is the case. The unsafety test requires the Court to decide whether, given what the Court now knows by the time of the appeal, it is sure that the jury would still have convicted. In lurking doubt appeals, the Court does not know anything that the jury did not know, i.e., there is no 'counterfactual' for the Court to consider. Indeed, it could be argued that the Court knows considerably less than the jury because it does not see or hear all the witnesses. The appellant is unable to point to any concrete reason why the jury might have made a mistake. If the Court was to begin to be more liberal in allowing appeals on the basis of lurking doubt, this could begin to undermine confidence in most convictions, as it is almost always theoretically plausible that the jury could have made a mistake. The Court must strike a balance between undermining the jury in this way, and seeking to correct injustice.

⁶²⁴ *ibid* at [14]

⁶²⁵ LH Leigh, 'Lurking Doubt and the Safety of Convictions' (2006) Crim LR 809.

⁶²⁶ *ibid*, 811-2.

⁶²⁷ *ibid*, 810.

4.9 Conclusion

The 'unsafety test', under which the Court of Appeal has now operated for twenty years, is a product of the Court's chequered history in dealing with miscarriages of justice. The 'unsafety test' appears to provide the Court with broad discretion, but the Court has provided relatively clear guidance as to when a conviction will be unsafe. It has been argued that the essence of the unsafety test is whether, if the jury had known what the Court knows by the time of the appeal, all things considered, could a different verdict have reasonably been delivered? This means that the Court must first decide whether the position is any different at the time of appeal than it was at the time of trial, and, if so, whether that means the verdict could have been different. Whilst this does entail a degree of discretion, the Court has explained what circumstances will lead to convictions becoming unsafe. The 'legal model' of judicial decision-making, therefore, could be said to stipulate that the Court should apply the test in this way, as explained by the Court's jurisprudence. The Court is not entirely unconstrained in how it deals with appeals against conviction, but the 'unsafety test' itself entails certain principles which must be followed. This understanding of the 'unsafety test' forms the basis of the 'legal variable' discussed in the remaining chapters of this thesis. If this is how the Court is designed to operate, and this is how it in fact operates, this is a counterweight to the factual and demographic variables used, and is a measure of the legal model.

The remainder of this thesis is concerned with the empirical analysis of the Court's decisions. The next two chapters explain methods employed in the analysis, and the variables collected from each appeal in the study. The significant limitations of these methods are expressed. These chapters are important because they explain the measures of impartiality, which will allow for evaluation of how successful this thesis has been as measuring that concept. As was discussed in Chapter 3, the approach adopted in this thesis is quantitative and positivistic, and seeks to adhere to a replication standard. It is explained in the next two chapters how this is achieved.

Chapter 5

Study Design, Data Collection and Methods of Analysis

Introduction

This chapter explains how impartiality has been captured, on the basis of its 'observable implications'. Data have been collected from Court of Appeal transcripts by a process of quantitative content analysis. These have then been converted into independent variables. As this study is an empirical analysis of the concept of impartiality, it must be asked whether, or how far, the measures are valid, reliable, and replicable. This thesis employs hypothesis testing in exploring whether the variables are associated with the decisions of the Court. This chapter explains what is meant by hypothesis testing, p-values and statistical significance, and how this helps to overcome some of the problems caused by the 'fundamental problem of causal inference'. It will be explained that this thesis is concerned with correlation, not causation, and the limitations of this approach.

The study conducted in this thesis concerns only murder and rape appeals. The decision to only include these offences is explained in this chapter. Summary statistics relating to the murder and rape appeals in the sample, and how these relate to variables in the study, is provided. Finally, the binary logistic regression analysis procedure is explained, and it is shown how this is the appropriate procedure to address the research question raised in Chapter 1.

5.1 Quantitative content analysis

In order to explore the Court's decision-making, the information in Court of Appeal judgments is converted into numbers (i.e., 'coded') for analysis. This is done following a process of a quantitative content analysis. Riffe, Lacy and Fico define quantitative content analysis as being the:

‘systematic and replicable examination of symbols of communication, which have been assigned numeric values according to valid measurement rules and the analysis of relationships involving those values using statistical methods ... to draw inferences’.⁶²⁸

Quantitative content analysis ‘[reduces] communication phenomena into manageable data (e.g. numbers)’ which can then be examined statistically.⁶²⁹ Three terms from the above definition are crucial to quantitative research: ‘systematic’, ‘replicable’, and ‘valid measurement’. This study has been designed to conform to these standards to allow for the analysis of the Court of Appeal’s decision-making. The extent to which this has been achieved is a key component of this thesis.

‘Systematic’ quantitative content analysis ‘requires identification of key terms or concepts involved in a phenomenon, specification of possible relationships amongst concepts, and generation of testable hypotheses regarding the potential relationships.’⁶³⁰ The key concept under analysis in this thesis is impartiality, which was defined and explained in Chapter 2. The ‘phenomenon’ under analysis in this thesis is that some appeals are allowed and some are dismissed. This thesis seeks to explore the relationship between independent variables as a measurement of whether the Court appeared to have decided appeals impartially. The data collection is systematic because a set of hypotheses have been developed in relation to the possible relationship between independent predictor variables and the outcome of appeals against conviction. The hypotheses and variables used in this study are fully explained in Chapter 6.

‘Replicability’ is an essential component of quantitative analysis. It requires an ‘exactness’ to the research definitions and operations so that later readers can fully understand what was done.⁶³¹ In relation to Empirical Legal Studies (ELS)

⁶²⁸ D Riffe, S Lacy and FG. Fico, *Analyzing Media Messages: Using Quantitative Content Analysis in Research* (2nd ed, Lawrence Erlbaum Associates 2005) 25.

⁶²⁹ *ibid*, 23.

⁶³⁰ *ibid*, 25.

⁶³¹ *ibid*, 26.

concerning judges, Epstein and King argued that ‘good empirical work adheres to the *replication standard*’, in that another researcher should be able to ‘understand, evaluate, build on, and reproduce the research without any additional information’.⁶³² It is only by following this standard, Epstein and King argue, that it is possible to know that the research is not biased and so can present knowledge about the court under observation.⁶³³

‘Valid measurement’ (or ‘validity’) in quantitative content analysis means that the data collected must accurately represent what is being measured.⁶³⁴ In relation to this study, this means that the measures used (the independent variables) must accurately capture the underlying concept of impartiality.⁶³⁵ Epstein and King argued that to produce reliable and valid inferences, researchers should, 1) invoke theories that produce *observable implications*, 2) extract as many implications as possible, and 3) delineate how they plan to observe those implications.⁶³⁶ As discussed in Chapter 1, the variables collected in this study do not completely capture the principle of impartiality; the measures do not, therefore, have full validity. This limits the strength of conclusions which can be drawn regarding the Court’s impartiality.

As Hall and Wright argued, content analysis appears particularly appropriate as an ELS methodology, because it resembles what lawyers and legal scholars already do.⁶³⁷ ‘Black-letter’ legal scholars frequently read a series of cases, collect information, and discuss their significance. Content analysis can bring a systematic rigour to the analysis of cases, which provides ‘a way of generating objective, falsifiable, and reproducible knowledge about what courts do and why they do it’.⁶³⁸ They argued that content analysis is more useful for some kinds of

⁶³² L Epstein and G King, ‘The Rules of Inference’ (2002) 69 U Chic L Rev 1, 27.

⁶³³ *ibid*, 31.

⁶³⁴ Riffe, Lacy, and Fico (n 628) 31.

⁶³⁵ Epstein and King (n 632) 62.

⁶³⁶ *ibid*, 47. (Emphasis added).

⁶³⁷ MA Hall and RF Wright, ‘Systematic Content Analysis of Judicial Opinions’ (2008) 96 Cali L Rev 63, 64.

⁶³⁸ *ibid*.

legal analyses than for others. They note four different uses of content analysis in empirical legal research:

- 1) projects investigating the bare outcomes of legal disputes,
- 2) projects investigating the legal principles of case outcomes,
- 3) projects investigating the facts and reasons that contribute to case outcomes, and
- 4) 'jurimetrics' that attempt to predict the impact of facts on litigation.⁶³⁹

They argue that content analysis can work well for the first three, but that jurimetrics overreaches the epistemological aims of content analysis.⁶⁴⁰

This study could be considered 'jurimetric', but it is important to note that there is no attempt to predict future decisions. It would best be considered research of category 3. As Hall and Wright say, their third category is suggestive of research which seeks to 'document trends in case law and the factors that appear important to case outcomes'.⁶⁴¹ Category 3 research can be contrasted with 'jurimetrics', which seeks to 'predict the likely outcome of litigation or appeals based on real-world or trial-record views of the facts'.⁶⁴² This thesis does not seek to predict the outcome of future litigation but seeks to discover which variables are 'predictors' of the outcomes of appeals which have already been decided. As Hall and Wright note, to predict future cases it would need to be assumed that the information provided in judgments is a complete reflection of everything which contributed to the decision. This is an assumption which is unlikely to hold.⁶⁴³

⁶³⁹ *ibid*, 85.

⁶⁴⁰ *ibid*.

⁶⁴¹ *ibid*, 91.

⁶⁴² *ibid*, 99.

⁶⁴³ *ibid*.

The quantitative content analysis of Court of Appeal judgments was the method of collecting the data. These data are designed to offer some measurement of the impartiality of the Court. The principles behind the measurement of is now considered. The particular variables designed to provide a measurement of impartiality are discussed in Chapter 6. The variables are also listed in Appendix A, readers who wish to review the variables earlier may want to turn to Appendix A.

5.2 Measuring impartiality

Epstein and Martin argued that conceptual questions can be answered indirectly by stating what the *observable implications* are of the concept which is being addressed.⁶⁴⁴ The observable implications of a theory are what would be expected to be seen in the data if the theory was true. The observable implications then help form hypotheses which are tested by the study. Stating what would be expected to be seen if the Court of Appeal was impartial allows impartiality to be tested by determining whether those expected observable implications did occur in the data collected. It is by ensuring that the variables under analysis derive from these hypotheses which test the normative question that one can ensure that research is theory-driven.⁶⁴⁵ By stipulating the observable implications of the concept, it is possible to identify objective measurements of the concept.

In order to understand how the observable implications of impartiality were developed it is helpful initially to see the design of social science research projects as following a process. It is by following a research design process that it is possible to ensure that the normative question can be addressed as closely as possible. Black suggested the following research design process:⁶⁴⁶

⁶⁴⁴ L Epstein and G King (n 628) 62.

⁶⁴⁵ See MT Croes 'Explaining the Dealings of Dutch SMEs with Potential Legal Problems: A Plea for a Theory-Driven Approach' (2013) *Int J L in Context* 239, 243.

⁶⁴⁶ TR Black, *Understanding Social Science Research* (2nd edition, SAGE Publishing 2002) 6.

1. State questions and hypotheses, identify variables
2. Determine design structure
3. Identify population and sample
4. Select statistical tests for assessing hypotheses
5. Carry out, plan, and collect data
6. Analyse, draw conclusions, and evaluate

The researcher can go back to previous stages as the research progresses, for instance to modify the research questions, or to formulate new, or modify, hypotheses and so on. The research process must first begin with an overall question or problem which calls for evaluation, which then becomes a specific research question. As stated in Chapter 1, the research question addressed in this thesis is whether the Court of Appeal appeared to have determined the sampled appeals in an impartial manner. This question leads to the development of hypotheses which are tested. There is an overall null hypothesis (H^0) that the Court of Appeal is impartial, and this is analysed using a series of null hypothesis tests and modelling. The alternative thesis hypothesis (H^1) is that the Court appeared to have determined appeals in a partial manner. The research design of this study means that it will not be possible to conclude that the Court lacked impartiality, as a finding of a lack of impartiality would require extraordinary evidence which is beyond the limits of this study. What can be tested is whether variables which are more indicative of impartiality or partiality show the strongest association with the outcome of appeals.

5.3 Impartiality and its ‘observable implications’

The selection of variables used in this study has been driven by the kinds of variables used in earlier studies of judicial decision-making. In Posner and de Figueiredo’s study entitled ‘*Is the International Court of Justice Biased?*’⁶⁴⁷ ‘bias’ was measured by assessing whether judges in the ICJ were more likely to vote in favour of countries similar to their own. They did this by categorising countries

⁶⁴⁷ EA Posner and MFP De Figueiredo ‘Is the International Court of Justice Biased?’ (2005) 34 J Legal Stud 599.

into blocs according to their relative region, wealth, culture, military and political alliances, and other such factors, to determine whether judges voted for countries in the same bloc as their home States.⁶⁴⁸ This acted as an imperfect but (they argued) suitable proxy for the concept of bias. Their observable implication of bias was that ‘a judge votes in an unbiased way if he or she is influenced only by relevant legal considerations...and not by legally irrelevant considerations’.⁶⁴⁹ ‘Legally irrelevant considerations’ include variables which are captured by the behavioural model, and includes some variables used in this thesis. For instance, the gender of the judge cannot be a legally relevant consideration. Such variables therefore may be an observable implication of a lack of impartiality.

In Voeten’s study of impartiality in the European Court of Human Rights,⁶⁵⁰ there were three observable implications of impartiality, which he termed ‘theoretically plausible sources of bias.’⁶⁵¹ These plausible sources of bias were cultural bias; bias incentivised by the judge’s career prospects; and personal policy preferences.⁶⁵² To capture these three theoretically plausible sources of bias, Voeten analysed data relating to, for instance, whether the judge originated from a common law or civil legal order; whether the judge’s State was formally socialist; and whether judges were expecting to retire at the end of the term.⁶⁵³ The hypotheses he tested were, for instance, that judges expected to retire shortly would be more likely to vote against their governments. He concluded that ‘the overall picture is mostly positive’ for the impartiality of the judges in that Court.

Numerous studies have shown that factual and demographic details appear to have an impact on outcomes. These are variables drawn primarily from behavioural or attitudinal research. Rachlinski and Wistrich concluded that

⁶⁴⁸ *ibid*, 602.

⁶⁴⁹ *ibid*, 600.

⁶⁵⁰ E Voeten ‘The Impartiality of International Judges: Evidence from the European Court of Human Rights’ (2008) 102 (4) *Am Pol Sci Rev* 417.

⁶⁵¹ *ibid*, 418.

⁶⁵² *ibid*.

⁶⁵³ *ibid*, 424.

behavioural factors appear to have a greater association with judicial decisions when that demographic is an issue in the case.⁶⁵⁴ For instance, Perisie found that female judges of the US federal appellate courts found for plaintiffs in sex discrimination and sex harassment cases more than male judges did.⁶⁵⁵ Moreover, she found an indirect effect of gender as male judges were more likely to find for plaintiffs when there was a female on the bench.⁶⁵⁶ Research conducted in Canada by Stribopoulos and Yahya suggests that gender may have more of an influence in certain fields of law.⁶⁵⁷ They found that:

In criminal cases involving sexual or domestic violence...there is a statistically significant tendency on the part of female judges to favour the interests of complainants and mothers. The converse of male judges voting in favour of the interests of accused persons and fathers is also true.⁶⁵⁸

It is theoretically plausible that for the offences of rape and murder, and especially rape, gender could be considered an issue. It is particularly pertinent, therefore, that gender is considered in this study.

Race / ethnicity has also been considered as a variable in American studies and found to be associated with particular decisions.⁶⁵⁹ Boyd found that female judges were more likely to find for plaintiffs in sex discrimination cases and black judges more likely to find for plaintiffs in race discrimination cases.⁶⁶⁰ Other behavioural variables include the religion of judges. Several studies have found that judges

⁶⁵⁴ JJ Rachlinski and AJ Wistrich, 'Judging the Judiciary by Numbers: Empirical Research on Judges' (2017) 13 *Annu Rev Law Soc Sci* (forthcoming).

⁶⁵⁵ JL Perisie, 'Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts' (2005) 114 *Yale LJ* 1759.

⁶⁵⁶ *ibid*, 1778.

⁶⁵⁷ J Stribopoulos and MA Yahya 'Does a Judge's Party of Appointment or Gender Matter to Case Outcomes? An Empirical Study of the Court of Appeal for Ontario' (2007) (45) *Osgoode Hall LJ* 315, 347.

⁶⁵⁸ *ibid*, 319.

⁶⁵⁹ AB Cox and TJ Miles 'Judging the Voting Rights Act' (2008) 108 *Columbia L Rev* 1.

⁶⁶⁰ CL Boyd, 'Representation on the Courts? The Effects of Trial Judges Sex and Race' (2016) 69 *Polit Res Q* 788.

holding certain beliefs determine appeals in certain ways.⁶⁶¹ For instance, Pinello found that variations in how judges decided cases in gay rights cases was associated with different religions.⁶⁶²

As was discussed in Chapter 3, the judicial studies branch of ELS is increasingly looking beyond behavioural variables, and so some scholars have tested whether legal factors are related to the outcome of cases.⁶⁶³ These variables have been tested over a variety of areas of law, such as immigration⁶⁶⁴ and crime.⁶⁶⁵ One study which utilised a large range of variables, including behavioural and legal variables, is the Sisk, Heise, and Morriss study. That study is similar in approach to the present study. They sought to determine whether the outcomes of appeals relating to a new sentencing rule varied depending upon a variety of different variables. They found that many factors, such as the gender of the judge,⁶⁶⁶ race of the judge,⁶⁶⁷ and his or her law school,⁶⁶⁸ were not associated with the outcome of cases. From this, they were able to conclude that ‘the law remains the alpha and omega of judicial decision-making.’⁶⁶⁹

Conversely, they found that several factors were statistically significant in explaining the outcomes of cases. They found that previous experience as a criminal defence lawyer increased the tendency of the judges to vote the law (which was tough on criminal defendants about to be sentenced) unconstitutional.⁶⁷⁰ They expressed the concern that ‘the attitudes developed in

⁶⁶¹ See SS Ulmer, ‘Social Background as an Indicator to the Votes of Supreme Court Justices in Criminal Cases: 1947 – 1956 Terms’ (1973) *Am J Pol Sci* 622; BH Bornstein and MK Miller, ‘Does a Judge’s Religion Influence Decision Making?’ (2009) 45 *Court Rev* 112.

⁶⁶² DR Pinello, *Gay Rights and American Law* (Cambridge University Press 2003).

⁶⁶³ G Sisk, M Heise and P Morriss, ‘Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning’ (1998) 37 *NYU L Rev* 1377; FB Cross ‘Decision-making in the US Circuit Courts of Appeals’ (2003) 91(6) *Cali L Rev* 1457. See also Pinello (n 663).

⁶⁶⁴ MS Williams and AO Law ‘Understanding Judicial Decision Making in Immigration Cases at the US Courts of Appeals’ (2012) 33(1) *Justice System Journal* 97.

⁶⁶⁵ See Sisk, Heise and Morriss (n 663).

⁶⁶⁶ *ibid*, 1451.

⁶⁶⁷ *ibid*, 1454.

⁶⁶⁸ *ibid*, 1460.

⁶⁶⁹ *ibid*, 1500.

⁶⁷⁰ *ibid*, 1470.

criminal [defence] practise appear to have persisted to the bench.⁶⁷¹ They also found that the criminal workload of the judge was a significant determinant of the outcomes of cases in their study.⁶⁷² The higher the workload of the judge, the less likely he or she was to find the law unconstitutional. They speculated that the reason for this might be that the sentencing guidelines under review would streamline the sentencing process, and would be a '[labour]-relieving measure' for the judges if the rules remained constitutional.⁶⁷³ This is heavily indicative of the managerial model of judicial decision-making, which was discussed in Chapter 3.

The Sisk, Heise, and Morris study used a range of variables indicative of a number of models of judicial decision-making. They used a wide range of personal background factors as variables. This study also utilises personal background factors as variables. For instance, the gender of the parties to the appeal, and the judges and lawyers are used as variables. The age of appellants, and complainants / the deceased are also used as variables. Ethnicity was considered as a variable, although it ultimately was not viable as a variable owing to difficulties in collecting the data from the judgments. Personal background variables are a measure of impartiality because they would appear to be legally irrelevant factors and so if they are associated with outcomes this would be more likely indicative of a lack of impartiality.

As is discussed below, however, it is not possible to conclude from this study whether the judges were 'influenced', as opposed to there simply being a relationship between variables and the outcome of appeals. Whilst background factors were used in this study, there are personal background characteristics which were not considered as variables. Further background variables could have included the previous experience of the judges, their educational background, personal interests and beliefs, and so on. These were not collected

⁶⁷¹ *ibid*, 1471.

⁶⁷² *ibid*, 1483.

⁶⁷³ *ibid*, 1485.

in this study because the intention was to use data from the transcripts only, whilst collecting this data would have required searching beyond transcripts. The reason for restricting the data collection to the transcripts was to ensure that the data source was reliable. The omission of such variables is a limitation of this study. Personal characteristic variables may be a good measure of impartiality, and so by omitting some the final models may be a less successful model of impartiality.

Cross attempted a comprehensive analysis of the US Courts of Appeals (as opposed to the US Supreme Court).⁶⁷⁴ He utilised the Songer database,⁶⁷⁵ containing data relating to several thousand reported decisions of the US Courts of Appeals. Cross considered the power of judicial ideology on outcomes, alongside personal characteristics, and a measure of the law. He found ideology to have less of an association with outcomes than seen in previous studies. At times, ideology was shown to have extremely limited value as a predictor of outcomes. He found that most factors explained only a small amount of variation in outcomes.⁶⁷⁶ He found that the explanatory power of the models improved when legal variables were analysed.⁶⁷⁷ Cross's study therefore added significant evidence that it is not personality, politics, or ideology which determines outcomes but judges following and applying the law. This thesis follows many of Cross's lessons, namely, that there is a measure of law, and appropriate cautions are expressed given the limitations of observational studies of courts.

As discussed in Chapter 3, since the 1970s American legal scholars and political scientists have analysed US Courts and decision-making. This has provided a well-developed literature on methodology, and large datasets, which scholars can utilise. The present study is not a study of an American Court, but rather is an assessment of the England and Wales Court of Appeal (Criminal Division). Studies of UK Courts or judges are relatively sparse and there is no similar study

⁶⁷⁴ FB Cross, *Decision Making in the US Courts of Appeals* (Stanford University Press 2007).

⁶⁷⁵ See <http://artsandsciences.sc.edu/poli/juri/appct.htm> <accessed 8 August 2016>.

⁶⁷⁶ Cross (n 674) Chapter 3.

⁶⁷⁷ *ibid*, Chapter 2.

on the Court of Appeal in this country. In recent years, there has been some studies of judicial (or quasi-judicial) decision-making in Britain, employing an ELS perspective. Thomas and Genn employed case simulation to explore decision-making in tribunals.⁶⁷⁸ They examined numerous factors, such as whether the tribunal was in paper form or an oral hearing; whether the members of the tribunal were legally qualified; and the impact of panel member background.⁶⁷⁹ They found that oral hearings were much more likely to lead to a successful outcome,⁶⁸⁰ and there was no statistically significant association between the background of the panel members (including their gender, age, household income, ethnicity, and religion) and their decision-making.⁶⁸¹ This study is noticeable because it may be said that it comes close to replication of experimental conditions, in that the same case was sent to different tribunals with certain features amended whilst the rest were held constant.

Cahill-O'Callaghan utilised the psychologist Shalom Schwartz's personal values model to determine whether judge's personal values influences their decisions.⁶⁸² She used this model to determine whether the personal values identified by Schwartz (self-direction, stimulation, hedonism, achievement, power, security, conformity, tradition, benevolence and universalism)⁶⁸³ were demonstrated by particular Supreme Court Justices in a hard case, *R (on the application of E) v JFS Governing Body*.⁶⁸⁴ She found that the judges in the majority demonstrated *universalism* while those in the minority demonstrated *tradition*.⁶⁸⁵ Thus, it appeared that the personal values of the judges in the *JFS* case did influence the legal decision.⁶⁸⁶ Cahill-O'Callaghan's work could be considered attitudinal in nature. Cahill-O'Callaghan discusses some of the studies of what she calls 'overt

⁶⁷⁸ C Thomas and H Genn, *Understanding Tribunal Decision-Making: A Foundational Empirical Study* (Nuffield Foundation 2013).

⁶⁷⁹ *ibid*, 3.

⁶⁸⁰ *ibid*, 8.

⁶⁸¹ *ibid*, 10.

⁶⁸² R.J. Cahill-O'Callaghan 'The Influence of Personal Values on Legal Judgments' (2013) 40(4) *J L & Soc'y* 596.

⁶⁸³ *ibid*, 603.

⁶⁸⁴ [2009] UKSC 15.

⁶⁸⁵ Cahill-O'Callaghan (n 682) at 610-1

⁶⁸⁶ *ibid*, 617.

characteristics,⁶⁸⁷ such as gender and ethnicity, some of which have been discussed above. She highlights the value in also considering personal values which she says can be tacit influences.⁶⁸⁸

Utilising the same Schwartz model, she later extended her previous study and examined the values expressed by individual Supreme Court Justices. She found that different judges did express different personal values,⁶⁸⁹ and those expressing the same values tended to reach the same decisions.⁶⁹⁰ It is unlikely that the research Cahill-O'Callaghan conducted on the UK Supreme Court could be conducted on the Court of Appeal (Criminal Division). This is because there is only ever a single judgment, making it impossible to determine which individual judge's 'personal values' are being expressed. It cannot be assumed (indeed, it is likely to be false) that the judge delivering the judgment is the only judge who had any input in crafting it. Thus, only 'overt' characteristics of the judges can be collected, but this is sufficient for the purposes of this analysis of which factors are, and are not, statistically significant predictors of the outcomes of appeals against conviction. This has the benefit that there is no attempt to impute particular values onto judges based upon what they say in judgments but the focus is upon objective measures and demographics.

Perhaps the single greatest presently available source of data on the decision-making of judges in Britain is the Crown Court Sentencing Survey. This was a data collection exercise administered by the Sentencing Council, in which judges in many thousands of criminal cases completed forms to indicate the factors taken into consideration when issuing a sentence.⁶⁹¹ Other information, such as gender, was also recorded. This resulted in large datasets, giving, importantly, the thoughts of the sentencing judge him or herself. Research based on the

⁶⁸⁷ RJ Cahill-O'Callaghan, 'Reframing the Judicial Diversity Debate: Personal Values and Tacit Diversity' (2015) 35(1) LS 1.

⁶⁸⁸ *ibid*, 10.

⁶⁸⁹ *ibid*, 18

⁶⁹⁰ *ibid*, 27.

⁶⁹¹ See <https://www.sentencingcouncil.org.uk/analysis-and-research/crown-court-sentencing-survey/> <10 September 2016>.

database is beginning to emerge. Pina-Sanchez and Linacre sought to explore the level of consistency in sentencing for assault cases.⁶⁹² They noted that 'consistency' entails that 'like cases are treated alike',⁶⁹³ and so is accordingly a principle similar to impartiality, conferring legitimacy and public confidence. They found that there was a substantial degree of consistency in sentence lengths across courts in England and Wales, contrary to concerns of inconsistent sentencing.⁶⁹⁴

Thus, it is submitted, the appetite for ELS studies of judges in Britain may be beginning to grow. This is to be welcomed, because users of courts and tribunals in Britain are far behind users of courts and tribunals in the US, when it comes to understanding how judges reach decisions. As was discussed in Chapter 4, there has been previous research on the decision-making of the Court of Appeal (Criminal Division), but this has focussed upon how well it performs in correcting miscarriages of justice. This thesis addresses an alternative question, regarding the relationship between a range of variables and the decision-making of judges.

5.4 The role of law in the decision-making process

In order to be able to make strong claims regarding the Court's decision-making, it would be beneficial if it could be shown that the variables analysed influenced, or *caused*, particular outcomes. For the reasons explained below, this cannot be shown by this study. Moreover, any allegation that factual and demographic factors, such as judicial attitudes or gender, influence outcomes, needs to surmount the claim that it was the law which determined the outcome. Judges are lawyers, well trained and experienced in applying the law. Judgments are usually framed in terms of the rules laid down by precedents or whether the court below correctly applied statutes. Lawyers make legal arguments to judges based on the law. If judges appear, and claim, to be deciding appeals based on the law applicable to the case, it must be evaluated whether this is reflected in the data.

⁶⁹² J Pina-Sanchez and R Linacre, 'Sentence Consistency in England and Wales: Evidence from the Crown Court Sentencing Survey' (2013) 53 *Brit J Criminol* 1118.

⁶⁹³ *ibid*, 1118.

⁶⁹⁴ *ibid*, 1128.

As was discussed in Chapter 3, more recent and advanced judicial studies research has sought to look beyond judicial attitudes and attributes, and towards the institutional situation of courts, including its approach to considering the rules of law. Moreover, the 'legal model' of judicial decision-making postulates that decisions are reached by impartially applying the law, and it is important that this model is tested.

As Cross says, the 'daunting' task of measuring ideology has been surmounted by researchers, but rarely has the task of measuring the law.⁶⁹⁵ The difficulty is caused by the unfeasibility of finding some measure of a whole body of law and then determining whether the case under analysis applied the law rather than ideology to reach a decision. In order to do this, one would need to determine in some objective manner what the law is. But this is made difficult by the fact that every case is different, and as such the law will be applied differently in different circumstances. This is further complicated by the question of how to determine whether the case under analysis was decided 'correctly' in light of the law. This makes the legal model difficult to frame as a falsifiable hypothesis.⁶⁹⁶ In contrast, the attitudinal model, for example, has developed a measure of ideology (the party of the appointing President), which can give rise to a falsifiable hypothesis. In Edwards's and Livermore's critique of attitudinal studies they point to the difficulty in coding precedent as being a significant pitfall.⁶⁹⁷ This is because each precedential case would need to be coded in some way to indicate its precedential value. This inevitably requires some interpretation of the cases, rather than an objective assessment.

Cross's study of the US Courts of Appeals represents one attempt to measure law in empirical studies.⁶⁹⁸ He sought to test how often US Courts of Appeals

⁶⁹⁵ FB Cross, *Decision Making in the US Courts of Appeals* (Stanford University Press 2007) Chapter 2.

⁶⁹⁶ FB Cross, 'Decision Making in the US Circuit Courts of Appeals' (2003) 91 *Cali L Rev* 1457, 1467.

⁶⁹⁷ HT Edwards and MA Livermore, 'Pitfalls of Empirical Studies that Attempt to Understand the Factors Affecting Appellate Decision Making' (2009) (58) (8) *Duke LJ* 1895, 1927-30.

⁶⁹⁸ Cross (n 695).

gave deference to certain trial courts' decisions. He found that judges' decisions which, according to the application of law, would be due more deference from appeal courts were indeed shown more deference. This, Cross argued, was a measurement of the law governing the case. He found that the law, as it was measured by him, significantly improved his ability to model outcomes. Furthermore, his measurement of the law consistently out-performed attitudinal or demographic factors as predictors of outcomes.⁶⁹⁹ This adds support to the claim that any attempt to model judicial behaviour which ignores a measure of the law is likely to be deficient.

In the study presented in this thesis, there is a measure of the law: the 'Did Error Occur?' variable. In Chapter 4 it was explained that the unsafety test requires the Court to consider whether it thinks the jury would still have convicted if they had known what the Court knows by the time of the appeal. It is only by being sure of this that convictions can have moral authority.

Frequently, appellants will argue that the law was wrongly applied by the lower court. The law governing the case includes the rules which state what the law is, in order to allow the Court of Appeal to decide whether it was correctly or incorrectly applied, and whether that makes the conviction unsafe. The 'Did Error Occur?' variable is coded 'yes' when the judges decide that an error occurred in the proceedings, requiring consideration of whether the error made the conviction unsafe. It is coded 'no' when the judges decided that no error occurred in the proceedings. This acts as a measurement of the law because in deciding whether an error occurred, the judges have to decide what the law is. If they decide that an error occurred in applying the law, the judges then have to decide what the effect of the error is. If this is how the judges determine appeals, it would suggest that the judges apply the unsafety test as interpreted by the Court. By including this variable in the analysis, it is possible to observe the relationship between the finding of an error and the outcome of appeals.

⁶⁹⁹ *ibid*, Chapter 2.

Whilst this measurement of law may mitigate some of the limitations of studies which do not measure law it is only partial, in the sense of being incomplete. It may be thought obvious that the Court will be more likely to quash convictions when an error occurred, and so this is bound to improve the accuracy of the models. Previous research on the Court of Appeal has shown that if the Court finds that no error occurred in the proceedings, appeals are unlikely to be successful. It is this finding which gives rise to the claim that the Court is restrictive due to being slow to quash convictions which only raise questions of fact. It is therefore an *a priori* hypothesis that the question of whether an error occurred will be a strong predictor of the outcome of appeals. All previous research indicates that this will be the case, and so will not be surprising to the legal community.

The better question may be not whether this variable is a predictor, but what interpretation can be given to the finding. Whilst previous research used this finding to suggest that the Court is not receptive to claims raising factual issues, the 'Did Error Occur?' variable is used in this thesis as a measurement of the law governing the case. This is because the 'unsafety test' has been interpreted by the Court of Appeal as meaning that appeals are likely to be successful when errors occurred, and unlikely to be successful when no error occurred. Whilst it may be questionable whether this is a suitable way for the Court to apply its powers, this is how the test has been applied. This variable, therefore, tests whether judges follow this interpretation of the test. Including a variable which captures the law is important for the capturing of impartiality. This is because the 'legal model' stipulates that decisions are reached by judges impartially applying the law. The extent to which that appears to be the case can only be assessed by including some measurement of the law relevant to the case.

5.5 The fundamental problem of causal inference

This study analysed the decision-making of the Court of Appeal through a quantitative content analysis of Court of Appeal decisions. Much social science research, and observational social science in particular, suffers from what is

known as the '*fundamental problem of causal inference*', which occurs because researchers can only observe the factual and not the counter-factual.⁷⁰⁰ This means that a researcher can only observe what happened, and what the demographic and factual details were in cases which were actually decided. It is not possible to see what actually would have happened if any of these factors had been different, and they cannot be controlled. Accordingly, it is not possible to say that a particular variable causes particular outcomes; it is correlation and not causation. The best way to impute causation is in randomised controlled experiments, in which everything except the variable of interest is held constant. Moreover, causation can only be reliably inferred if the same results have been demonstrated in replication studies.

This empirical study of the Court of Appeal using appeal judgments suffers from the fundamental problem of causal inference. This means that the research has been designed to be correlative, and not causal. Therefore, the analysis can show only association between variables and outcomes, not a cause and effect relationship.⁷⁰¹ The transcripts analysed in this study were not designed to be studied in this way, but were designed to provide answers to real appeals in the Court of Appeal. This highlights a further difficulty with correlational research: that the factors which in fact influenced a decision in one particular case might not have influenced the decision in any other case.⁷⁰² This means that, whilst correlational studies can be useful in locating patterns in the data, a great deal of evidence is needed before it can be suggested that there is a true substantive relationship.

Often in social science research the materials analysed were created directly for the research project. For instance, data is generated by surveys, interviews, or simulations, which will address, or the researcher hopes will address, the specific

⁷⁰⁰ L Epstein and AD Martin 'Quantitative Approaches to Empirical Legal Research' in P Cane and HM Kritzer (eds), *The Oxford Handbook of Empirical Legal Research*, (Oxford University Press 2010) 903.

⁷⁰¹ See C Thomas, 'How Should We Go About Jury Research in Scotland? A Response' (2016) Crim LR 915, 917.

⁷⁰² *ibid.*

question he or she is studying. In some kinds of studies, especially experiments or simulations, variables can be controlled, giving the possibility of drawing causal inferences.⁷⁰³ In empirical legal research of the kind conducted in this study, the judges giving judgment did not know their judgments would later be analysed statistically. The study is therefore observational and non-reactive. The judgments were delivered for the specific purpose of providing reasons for their decisions, to be read by lawyers, scholars, and other interested parties. They were not designed for the purpose of statistical analysis. It is for this reason that content analysis is necessary to extract the required data. This means there is no direct answer provided in the judgments to the question ‘Was this case decided impartially?’ or ‘Did the gender of the judge impact the outcome of this case?’ and so on.

This is what gives rise to the fundamental problem of causal inference; the problem that it is very difficult to design a study from which causal inferences can properly be drawn. In a true experiment, the researcher can control the possible causal variables. It is not possible in this study to run any kind of experiment to determine whether the judges were in fact impartial or whether other factors in fact *caused* particular outcomes, or what other factors influenced appeals. The research must be designed observationally by isolating the factual and demographic details of cases which have been decided, and by analysing them to determine whether certain factors appear to lead to statistical variations in outcomes.⁷⁰⁴ It is through the use of hypothesis testing that it is possible to calculate the chance that a particular variable has some effect on the outcome variable. The problem of causal inference inherent in the methods of this study means that this study is strictly concerned with the analysis of statistical association. It cannot show whether a variable *caused* a particular outcome, or whether judges were *influenced* by a particular factor. This method is useful, however, in explorative studies such as this one. As there have been very few previous studies of this nature in Britain, it is important to discover whether there

⁷⁰³ *ibid*, 918.

⁷⁰⁴ See C Winship and SL Morgan, ‘The Estimation of Causal Effects from Observational Data’ (1999) (25) *Ann Rev of Sociology* 659.

are statistical relationships between the variables analysed and the outcome of appeals.

5.6 Hypothesis testing, the p-value, and statistical significance

The purpose of null hypothesis statistical testing is to determine the likelihood that the null hypothesis, that is, that the variable has no relationship with the dependant variable, is true, by determining how far the data corresponds with what would be expected if the null hypothesis was true. This is established by calculating the degree and strength of any association between combinations of particular independent variables and the outcome of appeals. It can be analysed how likely it is that the null hypothesis is true by observing the p-value of each independent variable in a binary logistic regression analysis. A p-value is a percentage, between 0 and 1, which allows for some measurement of the strength of the predictive ability of the variables considered in the model of the Court's decisions. It is important to note that, for the reasons discussed below, the *smaller* the p-value, the *stronger* the association between the variables and the outcome of appeals.

In March 2016, American Statistical Association (ASA) issued a statement on statistical significance and p-values.⁷⁰⁵ The ASA stated that 'while the p-value can be a useful statistical measure, it is commonly misused and misinterpreted'.⁷⁰⁶ These concerns regarding p-values and significance testing were also raised by Nuzzo in 2014.⁷⁰⁷ The principles behind the p-value and significance testing are summarised by the ASA as so:

⁷⁰⁵ RL Wasserstein and NA Lazar, 'The ASA's Statement on P-values: Context, Process and Purpose' (American Statistical Association, 2016).

⁷⁰⁶ *ibid*, 7.

⁷⁰⁷ R Nuzzo 'Scientific Method: Statistical Errors' 506, *Nature*, 150.

‘A p-value provides one approach to summarizing the incompatibility between a particular set of data and a proposed model for the data ... The smaller the p-value, the greater the statistical incompatibility of the data with the null hypothesis, if the underlying assumptions used to calculate the p-value hold. This incompatibility can be interpreted as casting doubt on or providing evidence against the null hypothesis or the underlying assumptions’.⁷⁰⁸

There are a number of essential components to these principles which must be considered. The p-value summarises ‘the incompatibility between a particular set of data and a proposed model for the data’. What the p-value does not show is whether the null hypothesis is true, or the probability that random chance produced the data.⁷⁰⁹ The p-value shows the level of ‘statistical incompatibility of the data with the null hypotheses’. Independent variables with larger p-values suggest that the data is consistent with the null hypothesis and so the null hypothesis could be true, i.e. the predictor variable under consideration has no or limited measured relationship with the outcome variable. Smaller p-values can indicate that there is greater statistical incompatibility in the data than the null hypothesis would predict. Smaller p-values may therefore give some reason to doubt the null hypothesis.⁷¹⁰

The ASA statement noted that the concept of ‘statistical significance’ is an arbitrary figure. By convention, a p-value of lower than or equal to 0.05 is considered ‘statistically significant’, and a p-value of higher than 0.05 considered non-significant. This represents the 5% level of significance. The ASA was particularly critical of the position that null hypotheses are ‘rejected’ and alternative hypotheses ‘accepted’ if the p-value is statistically significant.⁷¹¹ The apparent flaw in this kind of reasoning is that if the null hypothesis is actually true but a variable is shown as statistically significant due to random chance, the person accepting the alternative hypothesis and rejecting the null is wrong. Therefore, a p-value of less than 0.05% means only ‘that the data are not very

⁷⁰⁸ RL Wasserstein and NA Lazar (n 705) 8.

⁷⁰⁹ *ibid.*, 9.

⁷¹⁰ *ibid.*

⁷¹¹ *ibid.*

close to what the statistical model (including the [null] hypothesis) predicted they should be',⁷¹² and a higher p-value 'indicates that the data are much closer' to the null hypothesis prediction.⁷¹³ Thus, whilst smaller p-values may allow for the null hypothesis to be rejected, there is still a chance that the decision to reject the null hypothesis is mistaken.

One further important point to note in relation to the understanding of p-values and statistical significance is that a larger p-value does not mean that the variable had no predictive power at all. As Greenland and colleagues noted, it is only if the p-value is exactly 1 that it could be stated that the variable has no predictive value or relationship at all.⁷¹⁴ This is because in calculating the p-value it is assumed that the null hypothesis is true. Any p-value below 1, however slight, means that the variable did have some association. This reiterates that p-values are not able to distinguish between actually true and actually false null hypotheses and alternative hypotheses. However, variables with larger p-values can show that the data only deviate slightly from what would have been predicted if the null hypothesis was true, and so a researcher may be more likely to be mistaken in rejecting a null hypothesis with a larger p-value.

Thus, the use of p-values and null hypothesis significance testing has difficulties. This thesis has utilised p-values and null hypothesis testing when considering the impartiality of the Court of Appeal. As the ASA statement has made clear, however, a statistically significant finding at the 5% level is only weak evidence against a null hypothesis. Whilst the ASA's view was that p-values provide weak inferential evidence against the null hypothesis, it did agree that p-values can be useful in summarising data.⁷¹⁵ Some contributors to the ASA statement noted that p-values are less controversial, and more useful, in explorative studies as

⁷¹² *ibid*, 8.

⁷¹³ *ibid*.

⁷¹⁴ S Greenland et al 'Statistical Tests, P-Values, Confidence Intervals, and Power: A Guide to Misinterpretations', a contribution to the ASA statement roundtable, supplemental submission No 21.

⁷¹⁵ RL Wasserstein and NA Lazar (n 705) 8.

summary statistics for sets of data.⁷¹⁶ P-values do, therefore, have value in a study such as the present. They provide a quantification of variables which have a stronger relationship with the dependant variable. This thesis recognises the arbitrary character of statistical significance testing and so has been careful to present all the data and p-values found in the statistical tests. It is important to note that one key limitation of this study is that a p-value must be carefully interpreted, and that a finding of 'statistical significant' is an inherently tenuous measure of whether a null hypothesis is true.

The results of the analysis conducted in this thesis is not contingent upon the p-values alone but also the effect they have on the ability to correctly classify successful and unsuccessful appeals against conviction. This thesis also uses summary statistics, classification tables, R^2 , and confidence intervals as measurements of how well this study has captured the principle of impartiality. It is important to note that the ASA did not state that p-values are invalid, but the statement was related to definitional issues and the amount of confidence researchers can have in relying upon them, and the accurate communication of statistical results. The primary concern of the ASA is that policy decisions are made, such as whether to continue with a clinical trial, based on statistical significance, and that publishers only tend to publish statistically significant findings.⁷¹⁷ Given that statistical significance is an arbitrary figure, it may not be appropriate to reach clinical decisions based upon statistical significance. Clearly, this issue does not arise in this study. Necessary caution has been expressed throughout this thesis about the results of this study, given the observational nature of the data collection which could not replicate experimental conditions. Despite the reservations recently issued by the ASA, null hypothesis significance testing is an important and useful method of determining the predictive ability of independent variables against dependent variables, as long as these limitations are observed and heeded.

⁷¹⁶ See DA Berry, 'P-Values Are Not What They're Cracked Up To Be' a contribution to the ASA statement roundtable, submission No 5; MJ Lew, 'Three Inferential Questions, Two Types of P-Value', submission 14.

⁷¹⁷ See RL Wasserstein and NA Lazar (n 705).

5.7.1 Parameters of sampled appeals

The cases analysed in this thesis are the whole population of available murder and rape appeals against conviction decided between January 2006 and December 2010. This is a sample of the total workload of the Court of Appeal. Each case was downloaded and read alongside a template. The template was designed to allow as many features of the case to be identified with a Yes / No answer, or for qualitative data to be coded into dichotomous items labelled as 0 and 1 for the purposes of subsequent analysis. This template is reproduced and discussed in Chapter 6. The variables were extracted from each case by marking appropriately the sections on the template. Once a case was read it was immediately coded into SPSS (IBM SPSS Statistics, v. 24 2016, IBM Inc.). Only cases where an appellant was appealing a murder and / or rape conviction were included in the study. The precise parameters of the scope of 'murder' and 'rape' needs some attention, this is discussed below.

There were 472 full appeals against conviction included in the final dataset, 241 murder appeals, and 231 rape appeals. The offences of murder and rape were chosen for specific reasons. It would have been possible to follow previous studies and read the first 300 judgments from one year, or all cases from one year, or a random sample of cases, or all cases from a number of years. The latter option was excluded as the numbers involved would quickly become unmanageable for this study. The decision was made to focus on specific offences in order to explore decision-making in those offences. Since previous studies have already provided sufficient detail on the general decision-making of the Court of Appeal, it was decided that focussing only on certain offences would offer new insights into the decision-making of the Court. This means that there is no attempt in this thesis to generalise the findings to other offences decided in the Court of Appeal. This provides opportunities for further replicative research on different offences.

The specific offences of murder and rape were carefully chosen. They are amongst the most serious offences known to the law, and they are both

indictable-only offences. This is important because it means that any murder and rape convictions which are appealed can only be appealed to the Court of Appeal. Offences which are tried in the Magistrates' Courts are appealable to the Crown Court.⁷¹⁸ By focusing on indictable-only offences it is possible to be sure that all murder and rape convictions appealed will be included in the sample.

Attempted murder and attempted rape were not included in the sample. This is because these are separate offences to completed murder and rape, and the intention was to keep the offences under analysis as homogenous as possible. Further, applications for permission to appeal were omitted from the sample. There are several reasons for this. The primary reason is that the outcome of appeals and applications for permission is different. In full appeals against conviction, the outcome is that the conviction is quashed or upheld, while in applications for leave to appeal the outcome is that the grounds of appeal are / are not reasonably arguable. When permission is refused this is equivalent to dismissing the appeal, but when permission is granted there is still a long way to go before the conviction is quashed. This meant it was difficult to subsume the applications for leave within the full appeals' dataset. As the dataset was already sufficiently large it was decided the problem could be avoided by simply omitting to include applications for leave to appeal in the sample.

It should also be noted that applications for permission to appeal are treated very differently; it is not comparing like with like. Often applicants will be unrepresented, and it frequently appears that renewed applications are add-ons in appeals against sentences. It is also much more difficult to extract data from renewed applications; transcripts are rarely longer than 4 or 5 pages. Finally, applications were excluded because it appears that most renewed applications are not available on the legal databases. This then raises the question of why some are available and some are not, potentially leading to biased data.

⁷¹⁸ Magistrates Court Act 1980, s 108.

As was explained in Chapter 4, age categories of complainants / the deceased have been mapped onto the *Sexual Offences Act 2003*. This is with the addition of a further category which is absent from the 2003 Act – the 16 – 17 age group. This is included as a separate group in order to separate adult complainants / deceased from those aged 16 and 17. As the sample size for the 16 – 17 age group is relatively small, it should not have a large impact on any further analysis. Table 5.1 shows the age profile of the complainants / deceased in the sample.

Table 5.1: Age profile of complainants / deceased in sample.

	Murder	Rape
Under 13	11	97
13 – under 16	4	29
16 – 17	15	19
18+	211	85

As will be discussed further in Chapter 6, relatively few deceased in murder appeals were children, whilst the majority of rape complainants were children. The effect of this is that the variables for age is more relevant to rape than murder. The following Table shows the cases in the sample categorised by gender.

Table 5.2: Appeals in sample by gender.

	Murder	Rape
Male	171	70
Female	19	212

As can be seen, there is a similar effect in murder appeals, in that one group is more common to one offence than the other offence. Namely, in the murder appeals, there are relatively few female deceased. The effect of this is that the variable which considers the association between the gender of the complainant / deceased and the outcome of appeals is more relevant to rape appeals as there is greater variation.

5.7.2 Parameters of sampled murder appeals against conviction

Murder is defined as the intentional unlawful killing of another human being.⁷¹⁹ Murder provides no particular definitional difficulties relevant to the data collection exercise, subject to the following point. If a person was charged with murder, but convicted of manslaughter this is excluded from study, simply because this is not a conviction for the offence of murder. Offences such as infanticide are excluded because this is not the same as murder.

A significant current issue within the law of murder relates to ‘joint enterprise’. Joint enterprise murder occurs when the principle offender (A) and the secondary offender (B) agree to commit crime A, but in the process of which (A) commits murder. The extent to which (B) is liable for the murder if he only foresaw that (A) may commit murder, rather than having intended that (A) commits murder, or encouraged him to do so, has been a problem area for the law of murder.⁷²⁰ It has generated many appeals which are included in this study. The case of *R v Jogee; Ruddock v R*⁷²¹ has recently stated that the law of joint enterprise murder had taken a ‘wrong turn’ in the case of *Chan Wing-Siu v R*.⁷²² *Chan Wing-Siu* stated as a matter of principle that foresight that (A) might commit murder could be sufficient for joint enterprise murder.⁷²³ *Jogee* decided that foresight may be evidence of encouragement, but would not be sufficient by itself for (B) to be guilty of murder.

The murder appeals included in this study may be heavily impacted by this conclusion of the Supreme Court. Of the 241 murder appeals in the sample, 109 (45%) were ‘joint enterprise’, in that one or more person was charged and / or convicted of the murder acting jointly. (Note that this does not mean all these appeals were joined appeals; it may be the case that only one member of the

⁷¹⁹ See BJ Baker, *Textbook of Criminal Law* (3rd edition, Sweet & Maxwell 2012) Chapter 11.

⁷²⁰ If (B) did intend or encourage (A) to commit murder the rules of accessory liability under the *Accessories and Abettors Act 1861* would likely apply and (B) would likely be guilty of murder.

⁷²¹ [2016] UKSC 8, [2016] UKPC 7.

⁷²² [1985] AC 168. This was followed, with some modification, on numerous occasions, see *Hui Chi-Ming v R* [1992] 1 AC 34; *R v Powell*; *R v English* [1999] 1 AC 1.

⁷²³ *Chan Wing-Siu*, *ibid*, at 175.

'joint enterprise' appealed). Only 15 (13%) of these 'joint enterprise' appeals were successful. The conclusion of the Supreme Court in *Jogee* that *Chan Wing-Siu* had resulted in an 'over-extension of the law of murder'⁷²⁴ may be reflected in the low number of joint enterprise murder convictions quashed. The Supreme Court in *Jogee*⁷²⁵ and the Court of Appeal in *R v Johnson and others*⁷²⁶ have intimated that it is unlikely appeals will be revisited in light of *Jogee*.

The question of whether the offence was said to have been committed by 'joint enterprise' is not used as a variable in this thesis. The reason for this is that joint enterprise in murder entails the specific issues discussed above which rarely apply in rape appeals. It may be possible in future work to broaden out the legal variables, in order to include the particular legal issues arising in individual offences. This is a limitation of this study, because the question of joint enterprise in murder is a significant issue in the law of murder, and following *Jogee*, is likely to become more important. The proportion of joint enterprise murder convictions quashed by the Court of Appeal is much lower than the overall proportion of murder convictions quashed.⁷²⁷ One plausible explanation is that this is what the law required, as the joint enterprise doctrine of foreseeability was firmly established in law,⁷²⁸ *stare decisis* would require appeals to be dismissed if the judge directed the jury correctly as to foreseeability. If this is the case, however, there were strong allegations that the law was unjust.⁷²⁹ A complete analysis of joint enterprise is beyond the scope of this thesis, which is intended as an exploration of the decision-making of the Court of Appeal in all murder and rape appeals. There is, however, scope for further empirical enquiry outside the scope and remit of this thesis.

⁷²⁴ *Jogee* [83].

⁷²⁵ *ibid*, [100]

⁷²⁶ [2016] EWCA Crim 1613.

⁷²⁷ See Chapter 7.

⁷²⁸ Especially by the House of Lords in *Powell*, see note 722 above.

⁷²⁹ See W Wilson and D Ormerod, 'Simply Harsh to Fairly Simple: Joint Enterprise Reform' (2015) Crim LR 3, and B Crewe, A Liebling, N Padfield, G Virgo 'Joint Enterprise: The Implications of an Unfair and Unclear Law' (2015) Crim LR 252.

Most of the appellants appealing a conviction for murder were convicted only of one count of murder, but 11 (4%) were convicted of more than one count. The number of counts of which an appellant was convicted, and whether an appellant was convicted of more than one offence of murder or rape, is a variable which is analysed in this study. Most appellants (79%) in murder appeals were represented by at least one Queen's Counsel, and this is also considered as a variable. 62% of murder appeals were heard by a judge of at least the rank of Lord Justice, sitting with two High Court judges.⁷³⁰ The corresponding figure for rape was 49%, meaning that more circuit judges, recorders and retirees were used in rape appeals. The ranks of the sitting judges are also variables utilised in this study.

5.7.3 Parameters of sampled rape appeals against conviction

Rape provides some substantial definitional problems, and this has impacted upon how this study has been conducted. The definition of rape for this thesis includes rape charged under section 1 of the *Sexual Offences Act 1956* and rape charged under section 1 of the *Sexual Offences Act 2003*, and rape of a child under the age of 13 charged under section 5 of the *Sexual Offences Act 2003*. Defendants are charged under the 1956 Act if the alleged offending occurred before 1 May 2004. The 2003 Act defines section 1 rape as being where (A) intentionally penetration of the vagina, anus, or mouth of another person (B), where (B) does not consent and (A) does not reasonably believe that (B) consents. Section 74 of the Act provides that consent means that the act was agreed to by choice, having the capacity to make the choice. The offence under section 5 of the 2003 Act does not require a lack of consent, and it is no defence that the appellant believed the complainant to be 13 or over.⁷³¹ As discussed below, this has obvious implications for the nature of the defence raised at trial. Section 1 of the 1956 Act defined rape as having sexual intercourse (vaginally or anally) with a person who does not consent and that the suspect was reckless as to whether the other consented.

⁷³⁰ Including Lord Chief Justice, Vice President of the Court of Appeal Criminal Division, and President of the High Court.

⁷³¹ See *R v G* [2009] 1 AC 92.

What may now be known as anal rape under section 1 of the 2003 Act was sometimes charged as 'buggery'; section 12 of the 1956 Act. This is included in the sample as it appears indistinguishable from what it now termed anal rape. Sexual offences relating to animals are omitted. Offences relating to photographs / images are omitted. In the rape appeals, 105 appellants (45%) were convicted of one or more offence of rape, and 166 appellants (71%) were convicted of more than one offence type. Multiple counts is included as a variable in this study. The outcome variable for this thesis relates to the rape conviction: so if the conviction for another offence is quashed but the rape conviction stands, this is recorded as an unsuccessful appeal. The same holds for murder if, for instance, a conviction for weapons offences is quashed but the murder conviction is upheld.

As was highlighted in Chapter 4, whilst there are only three sections which charge the specific offence of rape (2003 Act sections 1 and 5; 1956 Act section 1), there are numerous different kinds of rape. Historical sexual offences are often charged under the 1956 Act. It will be observed that the primary difference between the Acts is that what is now charged as oral rape may not have been rape under the 1956 Act, but may have been charged as indecent assault (section 14 of the 1956 Act), or some other offence. For the purposes of the data collection exercise the charging decision of the prosecutor was followed. This means that if what could now be charged as rape was charged as another offence under the 1956 Act, it is not included in the dataset as the appellant was not convicted of rape at the time. For the purposes of this thesis, it makes no difference that the appellant could now have been charged with rape.

The *Sexual Offences Act 2003* provides a number of definitional challenges. Its overlapping sections are well documented.⁷³² Other offences which could be considered as essentially identical to rape, such as section 2 assault by penetration, or section 9 sexual activity with a child, are not included. Again, the charging decision of the prosecutor will be adhered to, rather than stating what

⁷³² JR Spencer, 'The Sexual Offences Act 2003' (2004) Crim LR 347.

the appellant could have been charged with. The reason for this is that there could be important reasons why particular charges were placed; it would not be appropriate to begin second-guessing the charging decisions of the prosecutors.

Section 1 of the 2003 Act is silent as to any age requirement, whilst section 5 relates to children under the age of 13. Thus, offences relating to children aged 13 to 16, and offences against anybody over the age of 16 are all charged under section 1. This study has also utilised a variable for these categories, and for when the complainant was aged 16 or 17, or over the age of 18. There is a qualitative and quantitative difference between an offence committed against a person aged over 18 (i.e. an adult) and a child; an offence against a child aged under 16 (16 being the age at which consent can legally be given), and those aged 13 to under 16. These differences are highlighted by statistics relating to conviction rates for different age groups of complainants. Thomas found that different categories of rape had different conviction rates in Crown Court trials. Trials alleging rape of a female under the age of 13 resulted in conviction 58% of the time, but 75% of the time for rape of a male under 13.⁷³³ Rape of a female aged 16 or over had a conviction rate of 47%, whilst rape of a male 16 or over had a conviction rate of 58%. It is not clear whether it is the female complainant which leads to these differences, however, as rape of a female aged under 16 had a conviction rate of 62%, which is higher than the conviction rate for rape of a male under 16 (51%).

Table 5.1, above, provided the number of cases which featured the different age categories in the murder and rape appeals. Appeals against convictions for rape against complainants aged 13 – under 16, and 16 – 17 both had success rates of 42%, but note that the number of appeals for these categories was relatively small. The success rate for offences against children under the age of 13 was 35%, and 30% for adult complainants. These figures show some difference but this is not particularly large. The 42% success rate is likely to be impacted by the

⁷³³ C Thomas, *Are Juries Fair?* (Ministry of Justice Research Series 1/10, 2010) 32.

smaller sample size. The age category of the complainant is included as a variable in this study, and so will be returned to in Chapter 7.

An additional component which is analysed in this study is whether there is a relationship between the nature of the defence offered at trials, and the outcome of appeals against conviction. Again, the different categories of rape (by age) have an impact on this. Only a small number of appellants attempted to deny *mens rea* (consent or belief as to age) in offences against children under the age of 13 – this is obviously because this is not a defence to that offence. 89% of appellants denied the *actus reus* for this offence. As the age of the complainant increases, appellants were more likely to have denied *mens rea*: 37% denied *mens rea* in the 13 – under 16 group; 57% in the 16 – 17 group; and 63% in the over 18 group. Clearly, therefore, the nature of the offence has an impact upon the nature of the case at trial, and by implication the kinds of appeals which are likely to be successful in the Court of Appeal. However, as the success rate across the age categories are relatively similar (as discussed in the previous paragraph) this by itself does not appear to have a close relationship.

The inbuilt age categories, the division of complainants between genders, and the 1956 Act may lead to the conclusion that there is not only one offence of rape, but several. In this thesis, all these rape offences have been coded as one offence: rape. An alternative approach would have been to separate the different categories of rape into individual offences, in order to determine whether particular variables have relationships with outcomes in particular kinds of rape. Whilst this may be an avenue for further exploration and analysis in the future, this approach is not taken here. This is because the focus of this thesis is to explore the relationships between variables and the broad offences of rape and murder. To the extent that further refinement of the offence categories would have been beneficial, this is a limitation. However, all of the categories of rape offences are still broadly the offence of rape, and so using the single offence is justifiable.

5.8 Binary logistic regression

Regression analysis in its general form is a statistical procedure which seeks to model the value of an outcome variable by one or more independent predictor variables.⁷³⁴ In simple linear regression models, where there is a continuous outcome variable and one continuous independent variable, the regression equation takes the following form:

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon_i$$

This means that the outcome (Y) is modelled by β_0 , the point at which the line intersects the Y axis (i.e., the constant), and β_1 representing the slope of the line (the coefficient), and X representing the value of the independent variable, plus an error term. This is a simple (single independent variable) linear regression equation. This may be predicting a person's age by their height, or vice versa. In multiple regression models, there can be any number of independent variables, so the regression model would take the following form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon_i$$

The above equation may, for instance, be a model of house prices (Y) = the constant, plus the house footprint (B1, X1), plus number of windows (B2, X2), plus size of the garden (B3, X3), and so on (B_n, X_n). In the above formulas for simple and multiple linear regression, there is an assumption of a linear (i.e. along a straight line) relationship between the outcome and the independent variables and normally distributed continuous variables.

⁷³⁴ See A Field, *Discovering Statistics Using IBM SPSS Statistics* (4th ed, SAGE Publishing 2013), Chapter 8.

These assumptions are violated when considering the kinds of relationships and variables evaluated in this thesis. The outcome variable is not a continuous number (quantitative) but is a categorical binary outcome (there are two options; the appeal is successful or unsuccessful). As a result, binary logistic regression is appropriate. In binary logistic regression, rather than showing a linear relationship between the outcome variable and the independent variables, it seeks to predict the odds that a case will fall in a particular category given the values of the independent variables. The independent variables can be continuous quantitative data, categorical data, or a combination of both. The majority of the independent variables in this thesis are binary categorical variables generally in the form of an answer to the question: 'Did this variable apply? Yes or No'. The logistic binary regression formula is shown below:

$$P(Y) = \frac{1}{1 + e^{-(b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n)}}$$

Where $P(Y)$ is the predicted outcome of a case. It will be seen that the part of the model in brackets is identical to the multiple linear regression model shown above. The binary logistic regression equation allows the testing of whether any of the variables are statistically significant predictors of the outcome of appeals against conviction. Based upon the value of the combination of independent variables, each case is given a predicted value. The predicted value is represented by $P(Y)$, and is measured on a continuous scale between 0 and 1. A predicted value of 0.50 and below results in that case being predicted to have been a dismissed appeal; any predicted value above 0.50 results in the case being predicted to be an allowed appeal. It is then possible to compare the predicted values computed by the model with actual observed outcomes in the data in order to determine how well the model fits the data.

The binary logistic regression model also produces an 'odds ratio' (OR), which quantifies the strength of the association between each predictor variable and the outcome variable. The odds ratio is the chance of success with a characteristic present and the chance of success without that characteristic present. Variables which have no ability to predict outcomes will have an OR of 1 because the probability of success will be the same whether the variable is present or not. As explained above, this will be extremely rare because all variables will show some, however small, predictive ability and this will prevent ORs of 1. Any departures in either direction, to +ve or -ve infinity, indicates the strength of the predictive power of that independent variable and the outcome variable. If the confidence interval of an OR crosses 1, the variable will not be statistically significant because it would not be possible to know whether the true figure is 1. The further away from 1 the stronger the association; which in turn allows the p-value to be calculated. The further away from 1 the confidence interval range of OR is, the more likely it is to be statistically significant. An OR of greater than 1 indicates that the variable in question is associated with increased odds of being successful; lower than 1 indicates reduced odds of being successful.

As discussed above, p-values and statistical significance cannot explain overall how valuable the model is in predicting outcomes. It should also be recalled that this thesis does not seek to 'predict' future cases, but reference to 'predicting' is reference to the binary logistic regression model procedure which seeks to predict what the outcome of the appeals was.

In addition to p-values, further tests allow for greater scrutiny of the overall predictive power and fit of the binary logistic regression models. This is done by scrutinising various outputs: the Hosmer and Lemeshow goodness of fit test; classification tables; and pseudo R^2 statistics. The Hosmer and Lemeshow goodness of fit test evaluates how well the predicted outcomes match the actual observed outcome. Classification tables show which proportion of appeals against conviction were correctly classified as being either allowed or dismissed. There are two pseudo R^2 statistics: the Cox and Snell, and Nagelkerke. R^2

statistics present the amount of variation in outcomes which is explained by the model. The Cox and Snell is incapable of reaching 1 (i.e. 100% of variation explained by the model), and so is always the lower of the two, while the Nagelkerke is capable of reaching 1 and so is always higher. These two pseudo R^2 statistics represent an upper and lower bound respectively of the amount of variation in the outcome variable which is explained by the binary logistic regression model. This provides a measure of how many cases are correctly modelled.

These various measures of goodness of fit and predictive power of the models are utilised to quantify how strongly the variables predict outcomes. This relates to the difference between statistical importance and substantive importance. A variable could have statistical significance (i.e. have a low p-value) but if the overall fit or predictive power of the model is low it cannot be sustained that the variable in question plays any great role in decision-making. Conversely, if a variable has a higher p-value but good predictive power it may have higher substantive value. Thus, no particular measure can be used to determine the value of models or variables but the various measures should be considered together in a careful assessment of the models.

Binary logistic regression in SPSS provides the option of a 'forced entry' or a 'stepwise' method. In forced entry, all the selected variables, statistically significant or otherwise, are retained in the final model. The predictive power is then assessed for significance using the Wald statistic which can then be used to determine whether the pre-defined hypothesis (i.e. the null hypothesis) is accepted or rejected. There are two kinds of stepwise regression: backwards and forwards. In backwards stepwise binary logistic regression the model starts at step 1 with all the independent variables included in the model and then removes the variables one at a time based on their p-value and Wald statistics until the best fitting model is created. In forward stepwise regression, step 1 starts with no variables in the model and variables are included one at a time until adding further variables adds no improvement to the model.

Standard 'forced entry' binary logistic regression is used to construct the models in this thesis. Field notes that the alternative, either backwards or forwards stepwise logistic regression, is inappropriate for hypothesis testing as it has the goal of finding the most parsimonious model to fit the data.⁷³⁵ Croes is critical of the use of stepwise regression when the intention is to determine what kinds of factors have what kind of association with outcomes, rather than when one is seeking the most efficient or parsimonious model.⁷³⁶ Fitting the most efficient or parsimonious model is not the intention behind this analysis; but to find which variables are associated with outcomes and which are not. Once the models had been constructed, however, the model is checked using forwards and backwards methods to check for confounding or suppressing variables.

As an alternative to binary logistic regression, discriminant function analysis could have been used to evaluate the data. As is discussed in Chapter 6, the dependant variable used in this thesis is binary, but it may have been possible to code the dependant variable with four levels. If this had been done, discriminant function analysis may have been appropriate. Discriminant function analysis seeks to discriminate between two or more groups using predictor (discriminating) variables.⁷³⁷ Discriminant function analysis is best used when there are more than two groups to the outcome variable, rather than the binary allowed or dismissed categories used in the present study. Discriminant function analysis also has more exacting assumptions, such as an assumption that each group is drawn from a population which has a normal distribution.⁷³⁸ Such assumptions are violated in the present study and are not necessary for binary logistic regression analysis. For this reason, logistic regression analysis was selected as the most appropriate method of analysing the data, particularly where a large proportion of the considered variables are categorical in nature.

⁷³⁵ Field (n 734) 322-4.

⁷³⁶ MT Croes 'Explaining the Dealings of Dutch SMEs with Potential Legal Problems: A Plea for a Theory-Driven Approach' (2013) *Int J Law in Context* 239, 243.

⁷³⁷ Field (n 734) 654.

⁷³⁸ WR Klecka, *Discriminant Analysis* (Sage Publications 1980) 10.

In many areas in this thesis, the data collected from the cases is summarised in the form of percentages, frequencies, tables and graphs. This is useful for several reasons. Firstly, this thesis considers only murder and rape offences, meaning that any statistical information gathered is new information. The dissemination of this information in summary form helps to provide details about how such appeals are dealt with in the Court of Appeal. This helps to understand decision-making in the Court of Appeal. Secondly, the use of graphs and tables makes the findings more easily comprehensible to the non-statistical readership which is the majority of the audience of this thesis. Thirdly, summary statistics is primarily what has been provided in previous studies of the Court of Appeal. This thesis replicates in part previous research on the Court of Appeal. By following the same procedures it is easier to compare and contrast the offences of murder and rape with the findings of the earlier studies.

5.9 Conclusion

This chapter has explained the data collection exercise and the principles behind the statistical analysis conducted on the dataset of appeal cases. The intention is that, when read together with the next chapter, sufficient details have been provided to allow later replication of the study. The observational nature of the study has been discussed. This study could not have been designed so as to replicate experimental conditions. It is only in properly conducted and replicated randomised controlled trials that any kind of causation can begin to be inferred. Accordingly, the proper limits of this study should be understood. What is sought to be explored is whether the independent variables which are collected from each case are shown to be statistically significant predictors of successful appeals. The logic behind this is that if the Court was to be presented as having appeared impartial, certain variables, in particular the factual and demographic variables, should not be predictors of the outcomes of appeals in the Court of Appeal. This is because if a variable is a predictor of outcomes there is an implication of an association between the variable and the outcome.

The next chapter provides a complete list of the variables collected from each case, and evaluates the data collection template. Chapter 7 presents the findings of the study. This includes a descriptive analysis of the grounds of appeal argued, which replicates previous studies of the Court of Appeal, and the binary logistic regression analyses of the Court's decision-making. It is by carefully considering the outcomes of the next chapter that it will be possible to conclude whether the Court does appear to have determined appeals in an impartial manner.

Chapter 6

Evaluation of Variables and Methods of Study

Introduction

An important original contribution made by this thesis is the development of a measurement of the impartiality of the England and Wales Court of Appeal (Criminal Division), and the collection and coding of cases to create a dataset. In developing the dataset, many pieces of factual, demographic, and legal variables were collected from each case. It has been sought to capture the principle of impartiality by collecting a range of variables, and by testing whether relationships exist between variables and the outcome of appeals. As has been discussed previously in this thesis, how successful this has been is a key question, and this will be addressed in Chapter 8. By explaining the variables which measure the principle of impartiality, it can be examined how well the concept has been measured. In this chapter, full details of the variables included in the study and how they were collected and coded is provided. The data collected from the appeals includes factual and demographic details relating to the case, the judges and the appellants and complainants / deceased. The grounds of appeal have also been collected from each case. A legal variable which captures the Court's approach to dealing with grounds of appeal and the 'unsafety test' was also captured. This data collection exercise has sought to draw upon previous examples of quantitative judicial studies and Empirical Legal Studies (ELS), utilising the models discussed in Chapter 3.

The data from each appeal against conviction was collected utilising quantitative content analysis with the aid of a template. The template has been presented below, followed by an explanation of the decision-making process relating to the collection of the dependant variable and the independent variables. The effectiveness of the data collection process is evaluated in this chapter. Finally, this chapter summarises the strengths and limitations of the methods employed in this study. In particular, the limitations of this study are highlighted. It is important to highlight the strengths, but also acknowledge the limitations of the

study, in order to evaluate how well or closely the principle of impartiality has been captured, and so what sort of claims can be made about the impartiality of the Court of Appeal.

6.1.1 The dependant variable

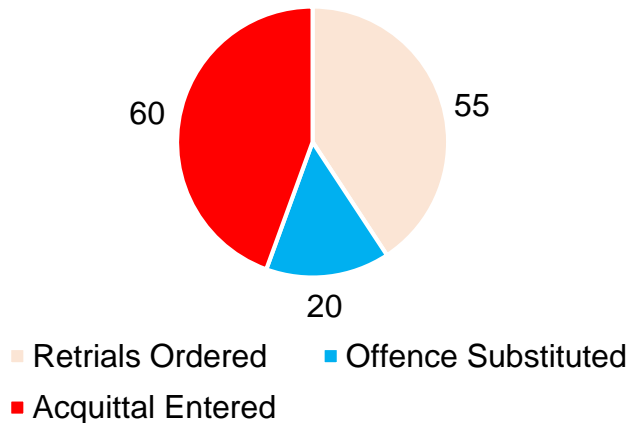
This is collected as an answer to the question: ‘Was the appeal successful?’ It is answered yes (coded 1) whenever a conviction for murder or rape was quashed, including when the appellant remains convicted of other offences. The dependent variable is coded 0 when the appeal was dismissed. An appeal is coded as being dismissed if the appellant’s appeal against either murder or rape was unsuccessful but appeals against other convictions were successful. This means that the dependant variable relates specifically to the murder or rape convictions. The outcome of appeals against conviction are considered to be binary in this thesis, and is coded accordingly.

This decision to code the outcome of appeals in a binary fashion could be challenged. This is because if an appeal is successful the Court has a number of options available to it. It can quash the conviction and enter an acquittal; it can order a retrial; or can substitute a conviction for an alternative offence. It could therefore be argued that there are four levels to the dependant variable, and not two, as so:

1. Appeal dismissed
2. Appeal allowed, acquittal entered
3. Appeal allowed, retrial ordered
4. Appeal allowed, offence substituted

Figure 6.1 shows the treatment by the Court of the 135 successful appeals in the sample.

Figure 6.1: Treatment of successful appeals.



Across the whole dataset of 472 appeals, therefore, 55 (11%) appeals had retrials ordered, 20 (4%) had offences substituted, 60 (13%) had convictions quashed and no further action, and 337 (72%) of appeals were dismissed. There was no variation in the proportion of successful rape and murder appeals that were ordered to be retried (50% of successful rape appeals were retried, and 50% of successful murder appeals were retried). Only two successful rape appeals had a conviction for an alternative offence substituted, whilst 18 murder appeals did. The obvious reason for this is that having a murder conviction quashed invites a conviction for manslaughter to be substituted, and this is often the case when the reason for the successful appeal is a finding of diminished responsibility.⁷³⁹

There was some variation in the post-success decisions relating to different kinds of appeal. For instance, only two of the 23 appeals (8%) which were successful following a CCRC reference were ordered to be retried, whilst 22 (39%) of the 59 appeals which were successful on the basis of fresh evidence were ordered to be retried. The reason for this discrepancy can be explained by the power of the Court of order a retrial. The Court can only order a retrial if it thinks it is in the interests of justice to do so.⁷⁴⁰ Owing to the nature of CCRC referrals the cases are often old and will have been appealed previously, meaning it is less likely to

⁷³⁹ See, e.g. *R v Erskine; R v Williams* [2009] EWCA Crim 1425.

⁷⁴⁰ Criminal Appeal Act 1968 (as amended) s. 7.

be in the interests of justice to retry the case, whilst the uncovering of fresh evidence may increase the impetus to have the issue retried.

Thus, there may be differing considerations for certain kinds of appeal as to which of the successful appeal outcomes to apply in the case. It would have been an alternative option to code the outcome of appeals with these four levels rather than two. The dependant variable has been coded as a binary outcome in this study because the question of whether to order a retrial, substitute a conviction, or enter an acquittal, comes after the decision to allow the appeal. The Court must first have decided that they are going to quash the conviction, and the question primarily considered in this study is whether the independent variables are associated with that decision. Dividing the dependant variable into four levels is an alternative that could be pursued in future research. As this is an alternative way of coding the dependant variable, it must be borne in mind that the conclusions are related to the binary outcome, and so some of the complexity of appeals is potentially obscured.

6.1.2 Factual and demographic variables

There follows a complete list of the independent variables included in the study. Note that there is no particular qualitative significance or order to the variables in this list. These variables in combination are the 'observable implications' impartiality, or a lack of it, which are employed in this thesis to incompletely measure the concept. Each variable has a null hypothesis that it is not a predictor of the outcome of appeals against conviction, and this is what is tested in the binary logistic regression analyses presented in Chapter 7. The process of null hypothesis testing was discussed in Chapter 5. It was noted that that the null hypothesis is essentially a hypothetical state in which the variable has no association with the outcome of appeals. It is unlikely that any of these null hypotheses will be true, and so the interest in this thesis is to explore and observe the strength of any association, in order to consider what this means about the impartiality of the Court.

- Date variables: Each day of the week and month of the year was an individual variable, coded 1 if the appeal was handed down on that day, 0 otherwise. Most appeals are *ex tempore* (see below). In non-*ex tempore* appeals the date is the date the appeal was handed down.
- *Ex tempore*: Coded 1 when all the indications from the case are that judgment was delivered immediately following the hearing, 0 otherwise.
- Joined Case: Coded 1 when the appeal involved more than one appellant, no 0 when the appeal concerned only one appellant.
- Multiple rape or murder: This was coded 1 if the appellant was convicted of multiple offences of either rape or murder.
- Convicted of other offences: Coded 1 if the appellant was convicted of other offences in addition to rape or murder.

These three variables relate to the number of persons involved in each appeal and the number of offences of which the appellant had been convicted. These variables are designed to reflect the overall level of criminality the appellants in the sample were convicted of being involved in.

- QC Appellant Only: Coded 1 when the appellant was represented by Queen's Counsel and the Crown was not represented by a QC, 0 otherwise.
- QC Crown Only: Coded 1 when the Crown was represented by Queen's Counsel and the appellant was not, coded 0 otherwise.

- Female Counsel Appellant: Coded 1 when the appellant was represented by female counsel, 0 otherwise.
- Female Counsel Crown: Coded 1 when the Crown was represented by female counsel, 0 otherwise.
- Female Judge Present: Coded 1 if there was a female judge on the bench, 0 otherwise.
- LCJ Present: Coded 1 if the Lord Chief Justice at the time was sitting as a judge on the case.
- VP Present: Coded 1 if the Vice President of the CACD was sitting as a judge on the case.
- President Present: Coded 1 if the President of the Queen's Bench Division of the High Court was sitting as a judge on the case.
- Circuit judge present: Coded 1 if any of the judges in the case was a circuit judge at the time of the appeal.
- Recorder Present: Coded 1 if any judge on the bench was a Recorder at the time of the appeal.
- Retired judge present: Coded 1 if any judge on the case was a retired judge, 0 otherwise.

- Trial judge Sat in CA: Coded 1 when the trial judge has sat in the Court of Appeal either very frequently in the past or in the recent past, in the sample of cases. It is also answered yes when the trial judge was an influential or famous judge, or held advanced positions.⁷⁴¹
- Individual Judges: There were 11 judges who heard at least 20 cases in the sample of cases. Each judge was a variable which was coded 1 if that judge sat on the case, No otherwise.
- White British Appellant: This states the ethnicity of the appellant. This variable was coded 1 if the appellant was White British, and 0 otherwise, according to any statement of ethnicity made in the judgment. As this information was rarely disclosed, the decision was made to also search press reports.⁷⁴²
- Sentence severity: It was hoped to examine whether particularly high or particularly low sentences were associated with the outcome of appeals. The sentencing for murder and rape are very different (murder generally much higher) so they could not be simply compared numerically. Instead, they were coded categorically by sentence severity. The categories are:

Rape:

0 – 60 months – Low

61 – 120 months – Medium

121 – 180 months – High

⁷⁴¹ Missing data complicated the collection of this variable. The variable was not always able to be collected as the trial judge was not always named, accordingly, this variable was not used in the study.

⁷⁴² This also did not prove to be sufficiently reliable, meaning data collection complicated this variable. Ultimately, missing data was such that this variable could not be utilised in this study.

181 + (including life sentences) – Very high

For IIP sentences the term is doubled to give the notional determinate sentence equivalent (as per the sentencing rules) for coding. Other sentences such as (rare) community orders, are low.

Murder:

Under 12 years – Low

13-20 years – Medium

21-29 years – High

30+ years (including whole life orders) – Very high

These were then coded as dummy variables with the variable coded 1 if it applied to that case, 0 otherwise.⁷⁴³ These categories could be challenged. It may have been possible to categorise these sentences differently, and alternative categories may lead to different results. There is an element of subjectivity within these categories. An alternative way of categorising the data would have been to attempt to map more clearly into the sentencing guidelines for the offences. The Sentencing Council guide provides starting points for sentencing rape,⁷⁴⁴ and the *Criminal Justice Act 2003* provides minimum term starting points for tariffs in murder.⁷⁴⁵ These starting points were not used as categories in this study because the starting points, especially in rape, overlap with each other. Moreover, since the sentencing judge's sentencing comments were not available, it was often not possible to know which category the judge placed the offence, which is the main determiner of the sentence.

⁷⁴³ Missing data complicated the collection of this variable. This is discussed further in Chapter 7.

⁷⁴⁴ Sentencing Council, *Sexual Offences Definitive Guideline* (Sentencing Council 2016) 10, 29.

⁷⁴⁵ Criminal Justice Act 2003, Schedule 21.

Accordingly, these categories were derived by first noting the sentence which was handed down before observing the spread of sentences and constructing the categories to be used in the study.

- Appellant Good Character stated: Coded 1 if the transcript refers to the positive good character of the appellant, no otherwise.
- Bad Character Stated: This was coded strictly according to whether the transcript explicitly stated that the appellant had bad character. If the transcript was silent this was coded 0. If the transcript stated that the appellant had any previous convictions, this was always coded 1. All convictions, however minor, were treated as bad character. However, given the nature of the offences it was rare that very minor offences were referred to by the judges.

The subjectivity needed in collecting the data for this variable should also be noted. As the data was derived from the information provided in the Court of Appeal transcripts, it is possible that relevant information was not provided.

- Denial of *actus reus*: This was coded 1 if the appellant denied committing the *actus reus* of the offence.
- Denial of *mens rea*: This was coded 1 only if the appellant admitted committing the *actus reus* but claimed either consent in rape appeals, or a defence in murder appeals. If the defence was complete denial it was answered No, because denial of *mens rea* was not the only defence.
- Historical offence: Coded 1 if the alleged offence occurred 10 or more years before the appellant was convicted of it, 0 otherwise. For murder this

is the date of the death. For a series of rape this is when the first instance alleged offending is said to have started.

- Appellant child: This variable is coded 1 if the appellant was under the age of 18 at the time of the alleged offence, or when the offending was said to have begun.
- Child deceased / complainant age range: This variable was coded into dummy variables, with each coded 1 if the deceased / complainant fell into that category, 0 if not. Note that for a series or multiple rapes the age counted is the age of the youngest complainant when the offending is alleged to have first started. In murder appeals, the age is the age of the deceased at death. As was discussed in Chapter 4, the age categories have been developed by mapping primarily onto the *Sexual Offences Act 2003*. The age categories are:

Age under 13

Age 13 – under 16

Age 16 – 17

Age 18+

- Drink or drugs: This variable was coded 1 if either the victim and/or appellant was drinking / taking drugs at the time of the alleged offence. It was coded 0 if there is no evidence of drink or drugs in the transcript. It was coded 1 if drugs were involved in a murder (e.g. drug gang related).
- Victim male: This variable was coded 1 if the victim was male, by implication if this was coded 0 the victim was female.

- Known victim: This variable was coded 1 if the victim / complainant was known to the appellant, 0 otherwise. The following kinds of relationship were always be coded as being 'known': family, partners, friends and friends of friends, those living in the same street or block, and work colleagues, members of same / rival gangs etc. This is unless it is otherwise stated in the judgment. Strangers are cases where the victim / complainant was completely unknown to the appellant before the day / night that the offence was alleged to have occurred.

From the above list, it can be observed that a range of factual and demographic variables are included in the study. Some variables can be considered to be behavioural in nature, in that they relate to personal characteristics of the judges, lawyers, and deceased / complainants. Some of the variables relate to how the case was run at trial, in order to determine whether decisions made a trial are associated with decisions on appeal. Some of the variables in the above list may be considered to have attitudinal implications, such that certain attitudes may underlay the variable. For instance, the bad character variable is designed to explore whether the judges might have a certain attitude towards appellants who have previous bad character and who are now appealing convictions.

These variables capture in an incomplete way the impartiality of the Court of Appeal. One question to be addressed in this thesis is how well this concept has been captured. The variables analysed are not all the variables which could have been collected. For instance, behavioural studies have explored a range of judicial background factors to test the theory that a person's social background influences their behaviour.⁷⁴⁶ There are some judicial background variables included in this list, such as their gender and their ranks, but there are other factors which could be considered, for instance their educational background; ethnicity; the area of law they practised in, and so on. This means that the variables used to measure impartiality in this thesis is omitting variables which

⁷⁴⁶ For example, TE George, 'From Judge to Justice: Social Background Theory and the Supreme Court' (2008) 86 NCL Rev 1333.

could be relevant. This means that the results of this study do not allow strong conclusions to be drawn regarding the impartiality of the Court of Appeal one way or the other. However, it must be recalled that the intention behind this study is to be explorative in nature, and to begin to observe any patterns in the data. It is not intended that this study will resolve the question of the impartiality of the Court. As will be discussed in Chapter 8, this means that the concept of impartiality has not been fully captured by the variables analysed, but further research using the dataset can help to capture it more fully.

6.1.3 Grounds of appeal raised

In addition to the factual and demographic variables detailed above, the legal arguments raised in the appeals were also collected from each case. In Chapter 7, the relationship between the grounds of appeal and the outcome of appeals is discussed. The grounds of appeal raised in appeals against conviction are clearly an important component of decision-making in the Court. As was explained in Chapter 4, the Court has developed a clear method of how it deals with appeals. For appeals raising procedural irregularities, the Court must determine whether the error occurred, and if so, determine whether the error means the jury might not have convicted if the error had not occurred. In fresh evidence appeals, the Court must broadly consider whether they think the fresh evidence renders the conviction unsafe, usually by asking whether the jury might not have convicted if they had heard the evidence. In 'lurking doubt' appeals, the Court is being invited to consider whether the circumstances of the conviction lead them to have a doubt.

The grounds of appeal were collected in order to analyse which grounds of appeal were most likely to be successful in murder and rape appeals. This follows previous studies of the Court by Roberts⁷⁴⁷ and Heaton.⁷⁴⁸ These variables are

⁷⁴⁷ S Roberts 'The Decision-Making Process of Appeals Against Conviction in the Court of Appeal (Criminal Division) (DPhil Thesis, London School of Economics and Political Science, 2009). (Hereafter Roberts (2009)).

⁷⁴⁸ SJ Heaton 'A Critical Evaluation of the Utility of Using Innocence as a Criterion in the Post Conviction Process' (DPhil Thesis, University of East Anglia, 2013). (Hereafter Heaton (2013)).

legal factors which could have an impact on the decision-making of the Court, however they are not included as individual variables in the regression models. There are several reasons for this. Firstly, if certain grounds of appeal are associated with particular outcomes, this would say little about whether the Court was partial or impartial. As discussed in Chapter 4, for some grounds of appeal if the Court finds that that error occurred, the conviction will be automatically unsafe. For instance, if the Court finds that the judge incorrectly refused an application to stay the trial due to an abuse of process, or finds that the judge should have discharged the jury or accepted an application of no case to answer, the conviction is necessarily unsafe. This is a logical requirement flowing from how the 'unsafety test' works, because if the judge should have discharged the jury or accepted a claim of no case to answer, the appellant should never have been convicted. If the Court finds that the trial was unfair it is highly likely that the conviction will be unsafe.

Other grounds of appeal, however, do not follow the same logic – indeed this is the core of the unsafety test. For instance the summing up and evidential discretion grounds will usually require the two-step decision-making process discussed in Chapter 4. This two step-process can be seen in the data which is presented in Chapter 7. This means that the grounds of appeal raised do not show that the Court is following the law or the legal model, because the question for the Court is what effect the grounds of appeal have upon the safety of the conviction. The grounds of appeal themselves would not be a useful measurement of impartiality because it would not mean that the Court followed the law if certain grounds of appeal were shown to be statistically associated with certain outcomes.

An alternative approach has been taken in this thesis to measuring the extent to which the Court follows the law and the 'unsafety test'. The 'Did Error Occur?' variable (discussed further below) broadly measures the relationship between the interpretation of the law and the outcome of appeals. This, it is submitted, is a

better approximation of the legal model, and so impartial decision-making, than including individual grounds of appeal.

This section provides definitions of the grounds of appeal raised and how they were coded. Each ground was coded 1 if it was raised as a ground of appeal, 0 if not.

- Fresh evidence raised: Coded 1 if the appellant raised fresh evidence.

Fresh evidence and how it is dealt with by the Court of Appeal is an important consideration in this thesis. Fresh evidence is not included as a variable to capture the impartiality of the Court because fresh evidence appeals raise specific issues. The relationship between fresh evidence and the outcome of appeals against conviction is discussed in depth in Chapter 7.

- Misdirection / defective summing up: Refers to all complaints relating to the summing up or directions provided by the judge. It includes complaints that the judge did not respond accurately to jury questions, or that the summing up omitted parts of evidence or certain directions. It also includes complaints that the summing up was generally defective, unbalanced or unfair.
- Judicial intervention: Relates to the specific claim that some in appropriate intervention by the trial judge made the conviction unsafe.
- Prosecution disclosure: this relates to any claim that evidence was not disclosed to the defence.

- Prosecution errors, not disclosure: relates to any alleged errors in procedure not relating to disclosure. For instance, late applications.
- Inconsistent verdicts: This was coded 1 whenever the appellant claimed that inconsistent verdicts make the conviction unsafe.
- Police irregularity: This was coded 1 when the appellant claimed some impropriety of irregularity by the police. For instance: breach of PACE, insufficient investigation.
- Jury irregularity: relates to claims that there was an irregularity with the jury – for instance the jury was biased due to the membership of the jury, or the jury was rushed into reaching a verdict, etc.
- Not able to mount defence: This was coded 1 when it was alleged that for any combination of reasons the appellant was unable to mount an effective defence and this makes the conviction unsafe.
- Claim of lawyer error: This was coded 1 when the appellant claimed that his legal team made errors at or before trial. This includes, for example, not calling certain witnesses, general incompetence or not making an application before the judge.
- Other: This captures procedural irregularities which are not captured in any above categories, for instance: problems with the indictment.

Certain grounds of appeal related to the exercise by the judge of his or her discretionary powers. These are:

- Misuse of evidential direction: This variable was coded 1 whenever it was alleged that the judge should, or should not, have allowed a piece of evidence to be admitted. It also includes decisions such as allowing cross examination.
- Refused severance: This was coded 1 if it was alleged that the judge should have severed the indictment.
- Abuse of process, stay, discharge: This was coded 1 if it was alleged that the judge should have stayed, or discharged the jury due to an abuse of process. This includes stays of proceedings due to delay, unfair trial, etc. Includes a claim for a discharge of the jury for any reason.
- Refused no case to answer: This was coded 1 if it was alleged that the judge should have accepted an application of no case to answer. Alternatively it is coded 1 if it was alleged that the judge should have stopped the trial due to NCTA, even if a submission of NCTA was not made.
- Unfair trial as ground of appeal: This was coded 1 when the appellant claims specifically that an unfair trial made the conviction unsafe, or breaches of Article 6 means the conviction is unsafe. It is coded 0 if the appellant refers to an unfair trial but does not specifically state that the trial was unfair.
- Insufficient evidence, lurking doubt, generally unsafe: This was coded 1 if the appellant claimed that the court should quash the conviction on the basis that there was insufficient evidence, a Lurking Doubt, or is Generally Unsafe.

6.1.4 Legal and institutional variables

This thesis includes a general measurement of the law governing appeals. As was discussed in Chapter 3, in recent years the study of judicial behaviour has increasingly sought to include legal variables, in order to study the Court more broadly as an institution. The grounds of appeal discussed above are legal factors but they are not included as individual variables. The law governing the case has been captured by the following variable:

- Did error occur? This is coded 1 if, in the opinion of the Court, an error occurred in the proceedings. Furthermore, it is coded 1 if, in the opinion of the Court, any error occurred even if it was not argued as a ground of appeal. An error means that the Court have to consider whether the error made the conviction unsafe. Accordingly, minor blips do not necessarily constitute an error for the purposes of this variable.

At the outset, there is strong evidence to suppose that the answer to this question will be a strong predictor of the outcome of appeals against conviction. Firstly, previous research has shown that if the Court finds that no error occurred in the proceedings appeals will usually be dismissed.⁷⁴⁹ Indeed, this is a core of the complaint relating to the allegedly restrictive approach the Court has adopted to exercising its powers. Additionally, previous research has shown that procedural irregularities are what usually leads the judges to deciding that convictions are unsafe.⁷⁵⁰ If the Court finds that an error did occur, there is a presumption that the conviction will be quashed, unless the Court is sure of the factual accuracy of the verdict.⁷⁵¹ Accordingly, the null hypothesis of this variable is that it is a predictor of the outcome of appeals. This variable could be considered an indicator of the 'legal model'. As was discussed in Chapter 2, the legal model entails a belief in impartial judges fairly applying the law. As such, this variable

⁷⁴⁹ See Roberts (2009) 149.

⁷⁵⁰ See *ibid*, and K Malleon, *The Royal Commission on Criminal Justice Review of the Appeal Process*, Research Study No 17 (HMSO 1993).

⁷⁵¹ See e.g., JR Spencer 'Quashing Convictions for Procedural Irregularities' (2007) Crim LR 835.

adds as a counterweight to the behavioural factual and demographic variables which were discussed above.

Other variables, in addition to the 'Did Error Occur?' variable, are designed to capture the broader behaviour of the Court, including its relationship with the CCRC and the application of its leave process.

- **CCRC:** This is coded 1 if the case is a CCRC referral, 0 if not. The CCRC variable captures the institutional relationship between the CCRC and the Court of Appeal. The CCRC case statistics shows that CCRC referrals have a 66% success rate in the Court of Appeal, in terms of the percentage of successful appeals.⁷⁵² Thus, if the other variables do not account for the variation in outcomes, it is highly likely that this variable will be a predictor of the outcome of appeals. This variable acts as an additional counterweight to the factual and demographic variables, as the CCRC referrals undergo a lengthy review process by the CCRC.
- **Leave granted by single judge:** This is coded 1 if leave was granted by the single judge, 0 otherwise. By implication, if the above variable and this variable are answered No, leave must have been granted by the Full Court or by the trial judge. This is a further potential institutional variable because it covers the relationship between the permission process in the Court of Appeal and the decision-making process. It may be supposed that appeals which were granted permission by the single judge are more likely to have obvious procedural irregularities because the single judge granted permission on the papers at the first instance. This variable may capture, therefore, the strength of the grounds of appeal and the extent of the irregularities.

⁷⁵² <https://ccrc.gov.uk/case-statistics/> (accessed 4/7/17).

- Verdict unanimous: This variable was coded 1 if it is stated that the decision of the jury was unanimous, 0 if it is stated that the decision was by majority. Due to a high level of missing data when this variable was originally collected, this variable was constructed on the assumption that any missing data was a unanimous verdict. The assumption behind this lies in the fact that 80% of convictions are unanimous.⁷⁵³ The problem with this assumption is that those with majority verdicts may be more likely to appeal. For that reason this variable must be considered with caution. This variable might not be considered an indicator of a lack of impartiality because the strength of the evidence against an appellant is an important part of the 'unsafety test'. One measure of the strength of the evidence is whether the whole jury were convinced of guilt. Instead, this variable acts as a measure of the extent to which the Court considers the strength of the jury's belief in guilt when determining whether a conviction is safe.
- Number of Cases cited: This variable was a count of the number of cases cited in the judgment. Each case is only counted once, and it includes any sub references (e.g. references to cases made by the trial judge, or another judge in a different case). It is also counted as a reference to a case when a certain principle named after a case is discussed (e.g. a *Lucas* direction).
- Offence Rape: This was coded 1 if the offence was rape. By implication No means the offence was murder. This is coded as a legal / institutional variable because these two particular offences raise particular issues which are relevant to appeals. For instance, in murder appeals it is usually apparent that an offence has actually occurred, whilst in rape appeals this is often the core issue. Thus, this variable tests whether particular legal elements of these offences are associated with outcomes.

⁷⁵³ See *Court Statistics (Quarterly) Jan – March 2014*, main tables, table 3.17. Available at <https://www.gov.uk/government/statistics/court-statistics-quarterly-january-to-march-2014> <accessed 5 July 2016>.

6.2 Data collection template

In order to ensure systematic, and so replicable, data collection, the appeals against convictions were analysed with the aid of a template. After consulting the earlier empirical research on judicial decision-making a list of potential 'observable implications' of impartiality was initiated. The observable implications were formulated to be answered in the form of a Yes / No question, in which the template was marked 'yes' if the variable applied to the case, and 'no' otherwise. A pilot study was conducted in order to determine what kind of information could be extracted from most appeals, with the minimum amount of subjectivity needed in the data collection. Based upon the pilot, previous experience of reading Court of Appeal judgments, and previous research, the list of independent variables (observable implications) was finalised. The template shown below was then created to ensure that each variable was collected methodologically from each appeal. The template used is shown below, in order to aid analysis of the data collection methods. The template and data collection is then evaluated.

Part 1: The Case

- 1) Case Name:
- 2) Official Citation:
- 3) Date of judgment:
- 4) Number of female judges:
- 5) Ranks:
- 6) Names:
- 7) Conviction Quashed? Yes: No:
- 8) Leave granted by:

Single Judge

Full Court

CCRC

Trial judge

- 9) Offence: Rape: Murder:

Counts:

- Number of cases cited:
- Counsel:
- Ex-Tempore:
- No. of grounds of appeal:
- Successful ground(s) of appeal:

Part 2: The Appellant

- 10) Age at appeal: Under 18:
- Adult 18-59:
- 60+:
- Not stated

11) Age at time of offence:

Under 18:

Adult 18-59:

60+:

Not stated:

12) Gender – (Assumed Male): Female:

13) Bad character: Yes: No: Not stated:

14) Good Character: Yes: No: Not stated:

15) Ethnicity:

16) White_Brit? Yes: No: Not stated:

Part 3: The Trial

17) Convicted by unanimous verdict? Yes: No: Doesn't state:

18) Trial Judge Gender: Male: Female: Doesn't state:

19) Trial judge name:

Part 3a: Rape

20) Rape sentence: Months.

21) Sentence: Doesn't state:

22) Sentence: Life: Years:

23) Sentence: Extended Sentence: Yes: No: Years:

Offence:

- 24) Historical Offence? Yes: No:
- 25) Nature: Known: Stranger: Doesn't state
- 26) Defence: Denial: Consent: Doesn't state
- 27) Familial? Yes: No:
- 28) Prostitute? Yes: No:
- 29) Drink / Drugs Yes: No:

Complainant

- 30) Gender: Single female:
- 31) Gender: Single Male:
- 32) Gender: Multi female:
- 33) Gender: Multi Male:
- 34) Gender: Both genders:
- 35) Victim / Complainant a child? Yes: No: Doesn't state
- a. All Children? Yes: No:
- b. Age of youngest C at time alleged offences is said to have begun:

Part 3b: Murder

Offence:

- 36) Historical Offence? Yes: No:
- 37) Min Term: Doesn't state:
- 38) Nature / context:
- Known:
- Random attack:
- Gang:
- Joint Enterprise:
- 39) Weapon: Gun:

Knife:

Other:

40)Details:

41)Defence: Denial

42)Defence: Cut throat

43)Defence: Accident:

44)Defence: Lack of intent:

45)Defence: Positive defence:

46)Drink / Drugs Yes: No:

Victim

2) Gender: Single female:

3) Gender: Single Male:

4) Gender: Multi female:

5) Gender: Multi Male:

6) Gender: Both genders:

7) Victim / Complainant a child? Yes: No:

a. All Children? Yes: No:

Part 4: Grounds of Appeal

Fresh Evidence

47)Fresh Evidence Raised? Yes: No:

48)Reference to R v Pendleton? Yes: No:

49)Fresh evidence admitted? Yes: No:

50)Interest of justice passed? Yes: No: Not Stated:

51)Why evidence not admitted?

52)Capable of Belief? Yes: No:

53)Ground for quashing conviction? Yes: No:

54) Nature of FE:

Lurking doubt

55) 'Lurking doubt' Raised / residual discretion / insufficient evidence / generally unsafe? Yes: No:

Procedural Irregularity Grounds:⁷⁵⁴

Misdirection / Defective / Unbalanced Summing Up: Yes:

Judicial intervention: Yes:

Non-disclosure of evidence: Yes:

Prosecution errors (not disclosure): Yes:

Inconsistent Verdicts: Yes:

Police Irregularity / Misconduct: Yes:

Jury irregularity claimed: Yes:

Biased tribunal: Yes:

Not able to mount effective defence: Yes:

Claim of lawyer error: Yes:

Unfair trial / breach of Article 6: Yes

Other:

56) Evidence wrongly included / excluded

Expert Evidence

Witness Testimony

Witness statements

PACE

a. S. 78

⁷⁵⁴ Grounds of appeal modified from, Roberts (2009) at 64. Some of the categories have been modified.

- b. Confessions
- ii) YJPOA 1999
 - a. S. 41
- iii) CJA 2003
 - a. Hearsay
 - b. Previous Convictions / bad character:

Other evidence wrongly included/excluded:

Refused application for severance: Yes:

Refused abuse of process / stay / discharge: Yes:

Refused claim No Case to Answer? Yes:

Permitted Questioning: Yes:

57) Did an error occur? Yes: No:

58) Did error make a difference?: Yes: No:

Fair trial

59) Fair trial / fairness / human rights considered? Yes: No:

Part 5a: Quashed Convictions:

Quashed because of Fresh Evidence?

Quashed because of Lurking Doubt?

Quashed because of procedural irregularity?

Jury might not have convicted / verdict may have been different:

Trial was unfair:

Retrial ordered: Yes: No:

Part 5b: Upheld Convictions:

Court's reason for dismissing appeal:

Error did not occur:

Error made no difference:

Trial was not unfair:

Strong prosecution evidence:

Fresh evidence not admitted:

Fresh evidence not capable of belief:

Fresh evidence made no difference to safety:

Jury would still have convicted:

No lurking doubt

No doubt as to safety / sure conviction is safe:

-- End of Template --

6.3.1 Evaluation of data collection

As each case was studied, tick marks were placed in the appropriate boxes. The options for each variable corresponds with the coding in SPSS, with yes answers coded as 1, no answers coded 0, and not stated coded as -99. Some fields have only one box, which was ticked if the appropriate answer was yes. In these circumstances, all the other options were presumed to be answered 'no' and were coded accordingly in SPSS. Variables which are not usually binary, such as the names of judges, were made binary. For instance, judges' variables became the answer to the question 'Did Judge X appear in the case? Yes / No'. By designing the template in this way, the data collection exercise was simplified and so less prone to the likelihood of human error. The benefit of attempting to restrict the independent variables to binary response variables was seen most clearly here. Since the majority of variables were 'yes / no' answers, the possibility of disagreement over the correct answer is low. Although there is always the possibility of coding error, when there is a large amount of data it does not cause significant problems. This is because any coding error is likely to be random and so such errors cancel themselves out over time.⁷⁵⁵

The settling of the final variables to be collected using the template was not an immediate process, but was developed whilst the data collection took place. It was only when the data collection began properly that it could be known whether the template was working as intended, whether more data could be collected, or whether there were difficulties with certain variables. The template provided above is the final version of the template, which superseded earlier versions. When it was found that the template was not working as intended, for instance it was realised that a variable was being poorly collected, the variable was amended and the data collection process had to be redone for that variable. As discussed below, this occurred in, for instance, the ethnicity variables.

⁷⁵⁵ The same point is made by A Juliano and SJ Schwab, 'The Sweep of Sexual Harassment Cases' (2001) (86) Cornell L Rev 548, 557.

The template was not used to collect all variables, but some variables were created only once all the data had been collected. For instance, whether the trial judge had sat in the Court of Appeal was used as an independent variable, but does not appear on the template. This was because until all the cases were read it could not be known whether the trial judge did sit in the Court of Appeal. Additionally, some of the data collected on the template was ultimately coded in a different way in SPSS. For instance, the age of the complainant in rape appeals was collected as a continuous number, but was then categorised in SPSS.

For many of the variables the data was available in every case. For instance, it was possible to see who the judges and counsel were; the genders of the judges, appellants, complainants and the deceased; the offences the appellant was convicted of; when the convictions and the events in question took place, and so on, for every case. Importantly, the outcome variable – whether the appeal was allowed or dismissed – was always able to be captured uniformly.⁷⁵⁶ This meant that many of the variables were collected for all cases in the dataset in a uniform manner. This means that overall the dataset was a clean dataset, with relatively little missing data or subjectivity needed. Variables which required greater subjectivity were discussed above. The selection of variables for analysis was designed to allow for the clean collection of the variables across all the cases.

As discussed above, the coding of some of the independent variables, and the dependant variable, could be challenged. Decisions had to be reached during the data collection process as to how particular variables would be coded in a way suitable to the study. This creates a limitation to the weight which can be given to the results of the study. The results must be interpreted in a cautious manner because different coding decisions may lead to different results. As an initial explorative study of the decision-making of the Court this is not a great difficulty because the primary purpose of this study is to observe whether patterns

⁷⁵⁶ As was discussed above, in this thesis appeals where *any* conviction for murder or rape was quashed was counted as being a success, so there were no questions regarding 'partly successful' appeals.

exists between the variables and outcomes in the way that they have been coded. However, it reduces the extent to which it can be said that the variables are a valid measurement of the Court's impartiality. An important sub-question within this thesis is how well the concept of impartiality has been measured. These weaknesses in the data collection template and the range of independent variables clearly mean that impartiality has not been completely measured. This makes it difficult to draw conclusions regarding the impartiality of the Court.

This is compounded by the existence of missing data. Missing data makes it more difficult to use those variables in the final analysis, meaning that the data is essentially lost. The ethnicity, sentencing and variables relating to the trial judge were particularly susceptible to missing data. Due to difficulty in collecting some of these variables they had to be omitted from the analysis. This means that the findings of this study are subject to omitted variable bias, as known variables are omitted from the analysis. Whether these omitted variables would have an association with the outcome of appeals is unknown. Further research could be conducted to seek to rectify some of these limitations by expanding upon the range of variables studied, or considering different coding options.

The use of two different offences presented both opportunities and challenges in the collection of the data. As there were only two offences to be considered, as opposed to the many different offences which are heard in the Criminal Division, it was easier to capture variables which would apply to both offences. For instance, both offences contain *actus reus* and *mens rea* elements which are frequently contested and so this could be used as a variable throughout the dataset. Both offences concerned drugs, fresh evidence, and male and female victims of different ages in some significant numbers. However, some potential variables did not apply equally to both offences, or applied only rarely to one offence. For instance, it would have been possible, and it was originally intended, to capture whether appeals were more, or less, likely to be successful if a gun or knife was used. But whilst weapons were frequently used in murder appeals, any kind of weapon was relatively rarely disclosed in rape appeals. Consequently,

the weapons variable had to be removed as a variable as it was not feasible to use in the study. Further research could be conducted using a larger sample of only murder appeals to consider this question further.

As was discussed in Chapter 5, relatively few murder victims were in the youngest age categories, and most were over 18. This means that the younger age variables disproportionately applied to rape offences. Thus, when the age of complainant / deceased variables are considered in Chapter 7, the variable is more clearly measuring age as a factor in rape appeals rather than rape *and* murder appeals. A similar point is true about the 'historical' variable. There were very few historical murder appeals, while historical rape appeals were relatively common. It may be questioned whether the variables therefore measure the Court's reaction to the factors in *both* murder and rape, or if particular variables are relevant to only murder *or* rape. One potential solution was to have two datasets, one for murder and one for rape. If this was done the datasets would have been too small and so further cases would need to have been coded. It was decided that the best solution was to continue with one dataset but to acknowledge that some variables are more relevant to one offence than others. Since the intention is not to predict individual murder or rape decisions, but to observe patterns in the data in an explorative way, this is not a significant difficulty.

The collection of the data is limited to the facts disclosed by the Court. It was sought to limit the data sources to the Court of Appeal judgments in order to ensure that the data source was reliable. It had to be assumed that the facts disclosed in the judgments were accurate, and not biased. This assumption may not necessarily hold. Furthermore, there may have been evidence which was relevant to the case, or which could have been used as a variable, which is not disclosed in the judgments. For instance, there may be a strong piece of evidence which is not disclosed in the transcript which leads the Court to be sure of the appellant's guilt or innocence, and so determinative of the appeal. Moreover, it is known that some important information, such as ethnicity, could

not be collected. If that piece of evidence was disclosed it may have been feasible to use as a variable across the dataset. Thus, if the research was to be designed ideally there would be access to all the pieces of evidence so that a more complete picture of each case could be gained. This was not plausible for the kind of research which was conducted.

This limitation is related to the correlative and non-reactive nature of the study. It is correlative because the study has sought to explore whether there exists relationships between certain features of the appeals (which cannot be controlled by the researcher) and the outcome of appeals. As not all the information which would have been desirable was available, there is a limit upon the kinds of conclusions which can be reached regarding the decision-making of the Court. Moreover, as the variables were designed to allow for an assessment of whether the Court appeared to have determined appeals in an impartial manner, the limited range of variables places a limit upon how well impartiality has been measured. It has been stressed previously in this thesis that impartiality has not been completely captured by the variables under analysis, and so the results of the study cannot be conclusive of whether the Court appeared impartial. A summary of the limitations of this study is provided below.

6.3.2 Evaluation of the collection of grounds of appeal

A brief evaluation of the collection of grounds of appeal is necessary, because it forms an important element of this thesis and previous studies. The collection of grounds of appeal was far more difficult than collecting the factual and demographic variables. The primary difficulty is that it is not always easy to place particular grounds of appeal into categories. Some analysis and interpretation is needed in some cases. Grounds of appeal are often combined or built on top of each other. A typical example is where an appellant claims that the judge should have excluded a piece of evidence, and contingent upon that incorrect decision the judge unfairly summed up the totality of the case to the jury. This may arise when the judge decides to admit the bad character of an appellant, or admits hearsay evidence. If the Court of Appeal decides the judge acted correctly, the

summing up argument may fall away, but the Court will frequently still consider the fairness of the summing up. This creates a difficulty in deciding what precisely the grounds of appeal argued by the appellant were and how they were dealt with by the Court.

Lurking doubt appeals could be seen as qualitatively important because they are said to raise the appellant's innocence, or the potential that the jury reached a mistake. The coding of lurking doubt appeals is difficult because the phrase 'lurking doubt' is rarely used in the Court. Consider the following typical example: the appellant argues that the judge incorrectly admitted hearsay evidence (say), and without that evidence the appellant argues there was insufficient evidence, or it was unfair, to convict. If there is 'insufficient evidence to convict' the argument would appear to be that the Court of Appeal should have a 'lurking doubt' that there was sufficient evidence to convict. In this thesis, this would have fallen within the 'misuse of judicial discretion' and 'insufficient evidence, generally unsafe, lurking doubt ground', but other researchers may differ.

There is an element of subjectivity in the collection of grounds of appeal. This is because the lawyers do not always present their cases with distinct and neatly-packaged grounds of appeal. The researcher then needs to determine what the grounds of appeal are. As was explained in Chapter 3, the job of the appellant's counsel is to persuade the Court in light of all the circumstances that the conviction is unsafe. This will frequently require the combining or layering of grounds of appeal in order to create a strong case. This in turn can make it difficult to 'deconstruct' arguments back into their individual parts. However, despite these difficulties, the fact that four decades of studies, including this study, as will be shown, have found similar results relating to most grounds of appeal suggests that grounds of appeal can be collected with some reasonable reliability.

6.4 Research strengths and limitations

There is no known publically available dataset of decisions of the Criminal Division. The lack of a pre-existing dataset relating to the Court of Appeal has both positive and negative consequences for this study. The positive aspects of having to develop a new dataset is that the dataset was created specifically for this research project with the independent variables carefully collected for use with this study. There was no need to second-guess the intentions behind the coding decisions of others, as would be the case if a pre-created dataset had been used. This provided a greater element of control over the research. This thesis has the added benefit that the variables were collected for the purpose of addressing the question of the impartiality of the Court of Appeal.

The final dataset in this study contained 472 appeals against conviction. All the cases in the dataset were full appeals; not applications or renewed applications. Some examples of ELS research have used datasets far larger than 472, while some are smaller. Many examples of ELS utilise pre-coded databases of cases. The US Supreme Court Database is the largest, a dataset of US Supreme Court decisions dating back to 1791.⁷⁵⁷ The size of the database allows researchers to utilise very large sample sizes, occasionally of several thousand cases.⁷⁵⁸ The University of South Carolina maintains the Judicial Research Initiative (JuRI), which contains numerous databases of American courts at numerous levels.⁷⁵⁹ This includes the ‘Songer’ database of US Courts of Appeals decisions.⁷⁶⁰ Cross’s study of the US Courts of Appeals utilised the Songer Database of several thousand cases.⁷⁶¹ Sunstein and his colleagues assembled their own database of almost 5000 non-unanimous reported Federal Courts of Appeals cases,⁷⁶² and the same database was utilised by Boyd and her colleagues in their

⁷⁵⁷ HJ Spaeth, L Epstein et al 2016 Supreme Court Database, Version 2016 release 1 <http://supremecourtdatabase.org>.

⁷⁵⁸ See e.g. L Epstein, WM Landes and RA Posner, ‘Are Even Unanimous Decisions in the United States Supreme Court Political?’ (2012) *Nw U L Rev* 699.

⁷⁵⁹ <http://artsandsciences.sc.edu/poli/juri/databases.htm> <accessed 26 July 16>.

⁷⁶⁰ <http://artsandsciences.sc.edu/poli/juri/appct.htm> <accessed 26 July 16>.

⁷⁶¹ FB Cross, ‘Decisionmaking in the US Circuit Courts of Appeals’ (2003) 91 *Cali L Rev* 1457.

⁷⁶² See CR Sunstein, LM Ellman and D Schkade ‘Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation’ (2004) (90) *Va L Rev* 301, 311. See also CR Sunstein, D Schkade, LM Ellman, A Sawicki, *Are Judges Political? An Empirical Analysis of the Federal Judiciary* (Brookings 2006).

study of the influence of gender on outcomes.⁷⁶³ Included as part of JuRI is a database of non-US High or Supreme Courts, including a database of UK House of Lords decisions decided between 1970 and 2002.⁷⁶⁴

Other studies have used smaller datasets which are comparable to the sample size used in this study. For instance, Sisk, Heise, and Morriss's study utilised a dataset of 294 cases.⁷⁶⁵ Two studies from outside of the United States utilised samples of 260⁷⁶⁶ and 507⁷⁶⁷ decisions respectively. The former was a study of the Israel Supreme Court, the latter the Spanish Supreme Court. Recently, the *Howard League for Penal Reform* analysed 147 judgments to assess the importance placed upon maturity in sentencing decisions.⁷⁶⁸ As can be seen, there is a wide disparity in the sample sizes of empirical studies of judicial decision-making. The sample size of 472 used in this thesis is large enough for the kinds of statistical analysis conducted on the dataset. As is discussed in Chapter 7, owing to the number of variables included in this study an initial univariate analysis was conducted in order to ensure that there were not too many variables relative to the sample size. This process ensures that the models are not overfitted, whilst allowing for a greater number of variables to be explored.

As far as possible, the dataset is a complete collection of all murder and rape appeals against conviction decided between January 2006 and December 2010. This is all available appeals handed down by the Court, and not only reported cases. This means that the dataset has good coverage of murder and rape appeals which helps to ensure that any findings are reliable and replicable. This

⁷⁶³ CL Boyd, L Epstein and AD Martin, 'Untangling the Causal Effects of Sex on Judging' (2010) *Am J Pol Sci* 389.

⁷⁶⁴ <http://artsandsciences.sc.edu/poli/juri/highcts.htm> <accessed 26 July 16>.

⁷⁶⁵ GS Sisk, M Heise and AP Morriss 'Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning' 1998 *NYU L Rev* 1377, 1408.

⁷⁶⁶ K Weinshall-Margel, 'Attitudinal and Neo-Institutional Models of Supreme Court Decision Making: An Empirical and Comparative Perspective from Israel' (2011) *JELS* 556.

⁷⁶⁷ N Garoupa, M Gili and F Gomez-Pomar, 'Political Influence and Career Judges: An Empirical Analysis of Administrative Review by the Spanish Supreme Court' (2012) *JELS* 795.

⁷⁶⁸ Howard League, *Judging Maturity: Exploring the Role of Maturity in the Sentencing of Young Adults* (Howard League 2017).

has the benefit of avoiding any kind of selection effect which could be caused if only reported cases were used.

It is possible that some murder or rape appeals will not have been available on any of the legal databases. This can happen for a number of reasons: judgments may be withheld due to confidentiality issues or awaiting retrials. It is for this reason that the time period 2006 to 2010 was utilised, as the appeal judgments are less likely to be embargoed and so more likely to be available. It is not possible to know how many murder or rape appeals are missing from the dataset. However, it is likely that the numbers missing will be relatively small, because the numbers of appeals in the sample is relatively large. Importantly, the dataset is large enough in its own right for the purposes of the statistical analysis.

Several different legal databases (Westlaw, Lexis Nexis, Casetrack and bailii.org) were carefully searched for murder and rape appeals during the period. The first three databases are professional subscription legal research tools, and so provide the best coverage of decisions from the Court of Appeal. Bailii.org is a public freely available website which provides a relatively small, yet still large, number of decisions from a range of courts and jurisdictions. This was primarily used as a final confirmation that as few cases as possible were omitted from the sample. Westlaw and Casetrack were the most useful databases for the collection of cases, because they contained the majority of the cases in the sample.⁷⁶⁹ It was soon discovered that some appeals appearing on one database did not appear on another. By searching all these databases it was possible to reduce to a minimum the number of potential missing cases in the dataset. Thus, the dataset is as complete as it can be.

The limitations inherent in this kind of quantitative empirical study have been discussed throughout this thesis. The remainder of this section draws these

⁷⁶⁹ The Casetrack service was discontinued in February 2017. Cases which appeared only on Casetrack may now be lost unless available elsewhere.

limitations together in order to summarise what kinds of conclusions can be reached regarding the decision-making of the Court. The most significant limitations of this study are that by its design it is correlative and not causal and the variables utilised to measure the concept of impartiality are incomplete and could be subject to disagreement. These two limitations in combination reduce how far it can be said that the research successfully captures the appearance of the impartiality of the Court of Appeal.

It was explained above that the range of variables used to measure impartiality must be incomplete. This is an inherent limitation within the process of quantification of a concept such as impartiality. As Epstein and Martin explained, in all measurement exercises 'everything about the object of study is lost except for the dimensions being studied'.⁷⁷⁰ For instance, measuring a person by only their height omits a great deal of information about that person. This is a natural consequence of seeking to quantify phenomena which occur in the social world. A range of variables have been selected for analysis in this study, but whichever variables were selected, it would only be possible to capture part of the operation of the Court of Appeal. It is important that the right dimensions are abstracted for the purpose of the analysis and to 'capture all the parts that are essential to our research question'.⁷⁷¹ It is questionable how far all the essential parts of judicial decision-making could be captured in order to fully address the question of the Court's impartiality. In Chapter 8, it is considered what this means for the results of this study, and for judicial studies in the future.

The variables used to measure impartiality were restricted to the information available on the Court transcripts. This was to try to keep the data collection uniform and the sources of data restricted to a reliable source. There were some potential independent variables which were considered but rejected for use in the study, and these are discussed below. It is in theory conceivable that uncollected,

⁷⁷⁰ L Epstein and AD Martin, *An Introduction to Empirical Legal Research* (Oxford University Press 2014) 46-7.

⁷⁷¹ *ibid*, 47.

and *uncollectable* factors, such as eye-colour, height, attractiveness, the weather, whether a judge is ill on the day, and so on, could make the judges in the Court of Appeal decide differently by affecting their mood or how they view the appellant, victim or complainant. Furthermore, as was explained in Chapter 2, psychological research suggests that all decision-making is inherently unreliable or subject to heuristics. In light of that, it must be accepted that there are potential variables of interest which had to be omitted from the study. Under ideal conditions it would be possible to control for all variables and reduce this omitted variable bias, but due to the observational nature of the study this is not possible. This is an understandable limitation of the research.

The research conducted in this thesis is not an experiment, and so such unstated matters cannot be controlled and will rarely be disclosed in Court of Appeal transcripts. This means they cannot be considered as variables in this thesis. This means that the correct statement of the parameters of this examination of the appearance of impartiality is that this study is an examination of whether the Court of Appeal appeared to have acted impartially, or not, *regarding the variables which have been selected for analysis*. While this is an important limitation and caveat, the variables under analysis are a legitimate source of study and are capable of being an appropriate measure of the appearance of impartiality. This is because factual and demographic variables such as those utilised in this study have been regularly studied in the context of determining the decision-making of judges, notwithstanding these limitations. Moreover, in their role as judges there are certain kinds of partiality which is unjustified,⁷⁷² and these are encapsulated by the variables which are utilised in this study.

Whilst it may be theoretically possible that decision-making could be influenced by (for instance), the judges' individual moods on a particular day; the particular idiosyncrasies of judges; the attractiveness of complainants or appellants; the

⁷⁷² O Raban, *Modern Legal Theory and Judicial Impartiality* (Glass House Press 2003) 2.

time of the day that the decision was made,⁷⁷³ and other such uncollectable factors, the size of the dataset compensates for this. As the dataset is relatively large, if such factors did occasionally influence judges, or one particular judge in the dataset, they would be cancelled out by the rest of the cases in the study. Accordingly, the results of this study are not as vulnerable to the concern that such factors could not be collected.

The positivistic nature of this study entails an assumption that it is possible to model mathematically and statistically the behaviour of judges deciding real cases. Given that it is known that some variables are not collectable, this may lead to the challenge that it is not possible to model the Court sufficiently robustly. It is argued that this is primarily a philosophical question, and does not have major importance when one is conducting a quantitative socio-legal research project of this nature. It is a philosophical question because there are inevitably missing variables in all social science research projects. If one doubts the ability to construct models in such circumstances one is doubting the ability to gain knowledge in any social science.⁷⁷⁴ As Hammersley says, some positivist assumptions, such as that nature exists, and that true knowledge can be gained, cannot be avoided if one is conducting social science research.⁷⁷⁵ It must, therefore, be believed that the variables which have been selected bare some relation to the concept which is being studied and their analysis will provide facts and knowledge about the Court of Appeal's decision-making.

In light of the above, if independent variables are shown to be significant predictors of success and are shown to enhance the predictive ability of the models, this does not mean that the Court lacked impartiality. This is because other variables which are not tested could have a greater impact. Similarly, if variables are shown to be poor predictors of successful appeals, this would not demonstrate that the Court is impartial, as it could have been partial with respect

⁷⁷³ See S Danziger, J Levav and L Avnaim-Pesso, 'Extraneous Factors in Judicial Decisions' (2011) 108 PNAS 6889.

⁷⁷⁴ See discussion by M Hammersley, *Methodology: Who Needs It?* (Sage 2011) 34-5.

⁷⁷⁵ *ibid.*

to other, untested, factors. Moreover, the judges may be adept at appearing impartial, while not actually being so. Further research would need to be conducted in the future to determine whether any factors untested in this thesis are predictors of outcomes. For instance, the Court could be observed and appellants could be asked to self-report their ethnicity for further analysis; or a qualitative study could be conducted by talking to judges. All of these represent possibilities for future research. However, it was not appropriate to take these steps as this research has been designed as an explorative quantitative analysis of the Court of Appeal transcripts.

The research conducted in this thesis is unique for the Court of Appeal. The research presented in this thesis represents a starting-point for further analysis of the Court. The difficulties with using statistical measures to seek to draw inferences about human behaviour need to be remembered. As such, future research would be needed to determine whether the results of this thesis can be replicated. It is only if similar results are discovered through replication in different datasets and with different variables that the results of this thesis can be confirmed. This thesis does replicate and confirm some previous research on the Court of Appeal. The findings discussed in Chapter 7 offer some support, and are supported by, the analyses of previous research. For instance, both Roberts and Heaton found that appeals are unlikely to be successful if there was no error in the proceedings, and this result is found in this study. This is important because the results of this thesis can be considered to be a significant but sensible advancement of knowledge about the Court's decision-making.

As was explained in Chapter 2, it may be impossible to know whether the judges were subjectively biased or lacking in impartiality. To be able to do this, would require the ability to observe what was happening in the judges' minds at the precise time in history that any judgments was delivered. Clearly, this is not possible. Accordingly, there is no allegation in this thesis of subjective impartiality. Although the eventual goal would be to learn what the judges were thinking, the best that can be done is to assess the appearance of impartiality. It

is therefore important to state that whatever the results of the analysis are, there is no claim that any particular judge lacked impartiality in particular cases over which they presided in respect of any particular appellant demographic or case characteristic. To reiterate this, the judges have been anonymised in this study. This thesis takes a more neutral standpoint than alleging an actual lack of impartiality. Namely, this thesis seeks to determine whether the independent variables, as collected, are predictors of appeal outcomes, and these variables were designed to allow for objective impartiality to be measured. Based upon this evidence, it must then be asked, as per the question in *Porter v Magill*,⁷⁷⁶ whether a fair-minded and informed observer, having considered the facts, would conclude that there was a real possibility that of a lack of impartiality.

6.5 Conclusion

Across the previous three chapters, the methodology, methods, data collection process, and independent variables have been fully stipulated. This has been necessary to attempt to adhere to the social science aim of systematic and replicable research. This chapter has provided a complete list of all the independent variables which will be analysed in the binary logistic regression analyses. The purpose behind this was to allow for scrutiny of the variables and to allow for replication. It has been explained that the collection of the variables was a dynamic process, requiring some trial and error. The result of this process, however, was a dataset containing a wealth of reliably collected data on the Court's decision-making which was suitable for statistical analysis.

The next chapter presents the analysis of the Court's decision-making. This includes descriptive analysis of the grounds of appeal, and a range of binary logistic regression analyses. Additional analyses are then conducted to seek to complete the evaluation of the Court.

⁷⁷⁶ [2002] 2 AC 357.

Chapter 7

Findings and Analysis

Introduction

This chapter presents the descriptive and binary logistic regression analyses of the England and Wales Court of Appeal (Criminal Division) cases in the sample. This chapter presents two kinds of analysis of the Court's decision-making. Firstly, there is a descriptive quantitative analysis of the grounds of appeal raised and the success rates of grounds of appeal. It replicates the analysis conducted in Malleson's,⁷⁷⁷ Roberts's⁷⁷⁸ and Heaton's⁷⁷⁹ studies. The aim of this analysis is to determine whether rape and murder appeals raise different kinds of grounds compared to other kinds of offences, and to provide further descriptive analysis of the kinds of issues raised in these appeals. The aim of this element of the research is to determine whether the findings of previous studies are further confirmed by this study.

In order to gain a further understanding of decision-making in the Court, and to make an original contribution to the literature on the Court and the understanding of how it operates, it is sought to determine whether a range of variables are associated with the outcome of appeals. This begins with a univariate 'purposive selection' procedure. The primary aim of this process is to explore which variables are statistically stronger predictors of the outcome of Court of Appeal cases. Following the purposive selection procedure, the remaining variables are analysed in a number of binary logistic regression analyses. This chapter presents and evaluates the results of these binary logistic regression analyses. Finally, these analyses may miss some nuances in the decision-making of the

⁷⁷⁷ K Malleson, *The Royal Commission on Criminal Justice Review of the Appeal Process*, Research Study No 17 (HMSO 1993).

⁷⁷⁸ S Roberts 'The Decision-Making Process of Appeal Against Conviction in the Court of Appeal (Criminal Division) (DPhil thesis, London School of Economics and Political Science, 2009). (Hereafter Roberts (2009))

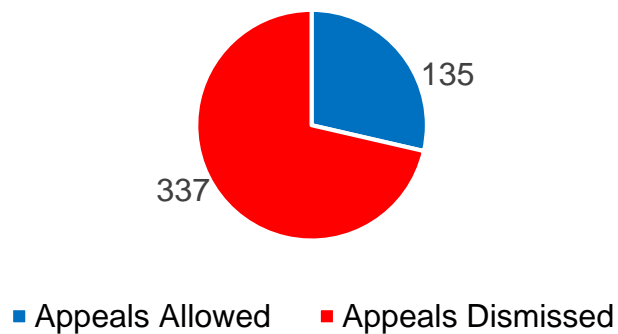
⁷⁷⁹ SJ Heaton 'A Critical Evaluation of the Utility of Using Innocence as a Criterion in the Post Conviction Process' (DPhil Thesis, University of East Anglia, 2013). (Hereafter Heaton (2013)).

Court. This is rectified by conducting some further analysis with varying datasets and variables.

7.1 Overview of sample profile

The final dataset contained 472 murder and rape appeals against conviction, decided between January 2006 and December 2010. Figure 7.1 shows the numbers of appeals which were successful and unsuccessful.

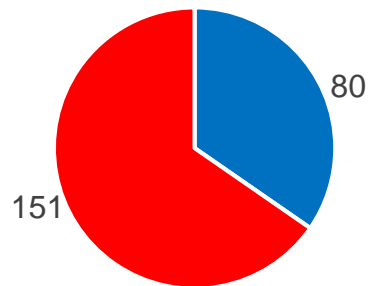
Figure 7.1: Outcome of appeals against conviction for Murder and rape 2006 - 2010.



The overall success rate was therefore 28%.

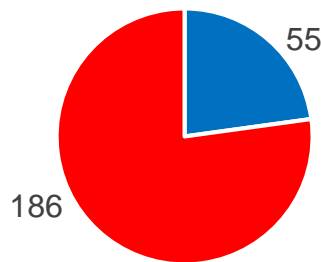
Figures 7.2 and 7.3, below, shows the success rates for rape and murder respectively.

Figure 7.2: Outcome of sampled rape appeals against conviction.



■ Appeals Allowed ■ Appeals Dismissed

Figure 7.3: Outcome of sampled murder appeals against conviction.



■ Appeals Allowed ■ Appeals Dismissed

The success rate for rape appeals was therefore 34%, and for murder 22%. The per-year overall success rate of the appeals in the sample is shown in Table 7.1.

Table 7.1: Outcome of sampled appeals per year, 2006-2010.

	2006	2007	2008	2009	2010
Appeals Dismissed	87	67	41	70	72
Appeals Allowed	33	24	25	21	32
% Successful	27%	26%	37%	23%	30%

Table 7.2: Percentage of murder and rape appeals successful per year, 2006 - 2010.

	Murder	Rape
2006	25%	29%
2007	20%	31%
2008	24%	51%
2009	16%	31%
2010	25%	36%

It can be seen that the percentage of murder and rape appeals allowed shows some fluctuation by year. It would be unwise to draw any particular conclusions regarding the attitude of the Court to murder and rape appeals from this simple table. It does not suggest that the Court has become steadily any more or less inclined to quash convictions as the years moved on. There is a noticeable increase in the success rate for rape in 2008, before a large decline in 2009. This is probably explained by the noticeable decrease in cases heard in 2008 whilst the number of successful appeals remained the same.

Data from official statistics show the following success rates for all appeals (i.e. not only murder and rape) decided between 2006 and 2010.⁷⁸⁰

Table 7.3: Overall success rates in Court of Appeal 2006-2010.

Year	2006	2007	2008	2009	2010
Success %	31%	37%	42%	38%	37%

⁷⁸⁰ Data collated from *Court Statistics Quarterly January to March 2014*. Additional tables, table 5.8 www.gov.uk/government/statistics/court-statistics-quarterly-january-to-march-2014. <accessed 16/09/2016>.

More recently, between 1 October 2014 and 30 September 2015, the overall success rate of appeals heard by the Court was 40%.⁷⁸¹ It will be observed that the overall success rate for murder and rape appeals is consistently below the overall success rate. The success rate in rape appeals always higher than the success rate in murder appeals. This may be considered as expected. In murder appeals there is rarely any dispute as to whether an offence was actually committed, while in rape this is often the core of the dispute. This means that there is one level of uncertainty in rape appeals which may not arise in murder appeals.

It may be possible to read this statistic as evidence of an insidious decision-making process of the Court. It may suggest that the judges are slower than usual to quash murder convictions than rape convictions. Murder is one of the most serious offences known to the criminal law, and so the relatively low success rate for murder might suggest the judges are particularly restrictive in allowing murder appeals. Based upon these statistics, it would again be unwise to draw this conclusion. Whilst it is a fact that the success rate is lower in murder than in rape, it is difficult to know whether this is due to a restrictive approach. What the difference in success rates does demonstrate, however, is that there are different success rates at appellate level for different offences. This should caution against drawing too many conclusions as to the decision-making of the Court of Appeal based on macro-level analyses of overall success rates, as there is high variability according to offence type.

7.2.1 Grounds of appeal raised in sample of murder and rape appeals

Roberts (2009),⁷⁸² and Heaton (2013)⁷⁸³ each analysed the grounds of appeal raised across their samples. As the grounds of appeal were not necessarily measured in exactly the same way, their results are not repeated in full. The results of this element of the research may not, therefore, be directly comparable

⁷⁸¹ *ibid.*

⁷⁸² See n 778 above.

⁷⁸³ See n 779 above.

to their results as grounds of appeal may have been coded or grouped differently. The intention behind this aspect of the study is to broadly replicate the previous studies and locate any obvious differences or similarities.

There are some important points of note about Roberts's and Heaton's findings relating to grounds of appeal. Both studies showed that summing up grounds of appeal were the most commonly argued and most likely to be successful.⁷⁸⁴ Summing up accounted for over 20% of the grounds of appeal in both studies and was successful the most often. The next most common grounds in both studies were claims that the judge had wrongly admitted or excluded evidence.⁷⁸⁵ Fresh evidence was the next most common. Fresh evidence represented 9% of the grounds of appeal in Heaton's study, and 6% in Roberts's study.⁷⁸⁶ In Heaton's study fresh evidence was successful on 17 occasions; nine occasions in Roberts's study. Lurking doubt was rarely seen in either study.⁷⁸⁷

Table 7.4(a) shows the number of times that particular grounds were raised across the whole sample of cases in the murder and rape appeals analysed in this study. Over the 472 cases, there were 885 grounds of appeal raised. The 'ground successful' column shows the number of cases in which that ground was one of, or the only, reason for allowing the appeal. Note that there are more successful grounds (154) than the number of successful appeals (135) because some appeals were successful for more than one reason.

⁷⁸⁴ See Roberts (2009) 64-6, Heaton (2013) 123-5.

⁷⁸⁵ *ibid.*

⁷⁸⁶ *ibid.*

⁷⁸⁷ *ibid.*

Table 7.4(a): Grounds of appeal raised in sampled appeals: Overall.

Ground of Appeal	Number of Cases Raised	Ground Successful (% success)	Success Rape / Murder
Summing Up	216	38 (17%)	23 / 15
Misuse of Evidential Discretion	166	24 (14%)	16 / 8
Fresh Evidence	130	59 (45%)	27 / 32
Refused No Case To Answer	51	2 (3%)	0 / 2
Unfair Trial Specifically	50	8 (16%)	6 / 2
Abuse of Process	44	0 (-)	0 / 0
Claim of Lawyer Error	35	3 (8%)	2 / 1
IE / LD / GU ⁷⁸⁸	30	3 (10%)	1 / 2
Inconsistent Verdicts	25	3 (12%)	3 / 0
Prosecution Error (Not Disclosure)	24	1 (4%)	1 / 0
Jury Irregularity (bias etc.)	20	3 (15%)	2 / 1
Prosecution Disclosure	18	2 (11%)	2 / 0
Judicial Intervention	12	1 (8%)	1 / 0
Refused Severance	7	0 (-)	0 / 0
Police Irregularity	6	1 (16%)	0 / 1
Not Able to Mount Defence	3	0 (-)	0 / 0
Others	48	6 (12%)	6 / 0
TOTALS	885	154	154

To allow for further comparison of the grounds of appeal raised in the individual offences, Tables 7.4(b) and 7.4(c) show the same data for rape and murder respectively. The percentages in brackets shows the percentage of rape and murder appeals respectively which raised each ground of appeal.

⁷⁸⁸ This stands for Insufficient Evidence/Lurking Doubt/Generally Unsafe.

Table 7.4(b): Grounds of appeal raised in sampled appeals: Rape.

Ground of Appeal	Number of Cases Raised	Ground Successful
Summing Up	106 (45%)	23
Misuse of Evidential Discretion	79 (34%)	16
Fresh Evidence	54 (23%)	27
Refused No Case To Answer	17 (7%)	0
Unfair Trial Specifically	23 (10%)	6
Abuse of Process	24 (10%)	0
Claim of Lawyer Error	22 (9%)	2
IE / LD / GU	13 (5%)	1
Inconsistent Verdicts	21 (9%)	3
Prosecution Error (Not Disclosure)	10 (4%)	1
Jury Irregularity (bias etc.)	11 (4%)	2
Prosecution Disclosure	4 (1%)	2
Judicial Intervention	9 (3%)	1
Refused Severance	5 (2%)	0
Police Irregularity	1 (1%)	0
Not Able to Mount Defence	3 (1%)	0
Other (indictments, etc.)	20 (8%)	6
TOTALS	422	90

Table 7.4(c) Grounds of appeal raised in sampled appeals: Murder

Ground of Appeal	Number of Cases Raised	Ground Successful
Summing Up	110 (45%)	15
Misuse of Evidential Discretion	87 (36%)	8
Fresh Evidence	76 (31%)	32
Refused No Case To Answer	34 (14%)	2
Unfair Trial Specifically	27 (11%)	2
Abuse of Process	20 (8%)	0
Claim of Lawyer Error	13 (5%)	1
IE / LD / GU	17 (7%)	2
Inconsistent Verdicts	4 (1%)	0
Prosecution Error (Not Disclosure)	14 (5%)	0
Jury Irregularity (bias etc.)	9 (3%)	1
Prosecution Disclosure	14 (5%)	0
Judicial Intervention	3 (1%)	0
Refused Severance	2 (1%)	0
Police Irregularity	5 (2%)	1
Not Able to Mount Defence	0 (0%)	0
Other	28 (11%)	0
TOTALS	463	64

The findings of this element of the research broadly confirm the previous studies, but there are some important differences. Like Roberts's and Heaton's findings, the three most commonly argued grounds of appeal relate to claims of errors in summing up, claims of misuse of judicial evidential discretion, and fresh evidence. When comparing the grounds of appeal raised in the murder and rape appeals, it will be seen that on the whole they raised broadly similar grounds of appeal. There are, however, a number of noticeable differences. Fresh evidence; claims that the judge should have accepted a submission of no case to answer; and allegations of defects with disclosure were raised more frequently in murder appeals, while inconsistent verdicts was more commonly argued in rape appeals. It will be observed that fresh evidence stands out as the most commonly successful ground of appeal in murder appeals. The summing up and misuse of evidential discretion grounds were less likely to be a successful ground in murder appeals than rape appeals.

7.2.2 Appeals raising fresh evidence as a ground of appeal

The statistics relating to fresh evidence are quite different in this study compared to the previous studies. Fresh evidence was raised in 130 cases in this study, this was 27% of all cases and 14% of all grounds of appeal. This latter figure was 9% and 6% in Heaton's and Roberts's studies respectively.⁷⁸⁹ In this study, fresh evidence was successful 45% of the times it was raised, which is significantly higher than the success rate in Roberts's and Heaton's studies (25% and 31% respectively).⁷⁹⁰ One noticeable observation from this study is that fresh evidence is actually the ground of appeal most frequently successful in both the murder and rape appeals. In the previous studies, the procedural irregularity grounds were most frequently successful.⁷⁹¹ One conclusion can safely be drawn from this statistic: fresh evidence is more likely to be raised and to be a successful ground of appeal in murder and rape cases than it is in the Court of Appeal generally.

Fresh evidence was raised in 76 murder appeals and 54 rape appeals. Whilst fresh evidence is raised more frequently in murder appeals, it was marginally more frequently successful in rape appeals. 27 convictions for rape were quashed on the basis of fresh evidence: 50% of the times it was raised. There were 32 murder convictions quashed on basis of fresh evidence, which is 42% of the times it was raised. This suggests that while fresh evidence is less forthcoming in rape appeals it is more often considered persuasive than fresh evidence in murder appeals.

For the purposes of this study, fresh evidence was divided into a number of categories. These were:

⁷⁸⁹ Heaton (2013) 125, Roberts (2009) 64

⁷⁹⁰ *ibid.*

⁷⁹¹ *ibid.*

- 1) Fresh expert evidence not relating to psychiatric evidence – for instance medical evidence of the cause of death or sexual injury
- 2) Fresh psychiatric evidence
- 3) Fresh statement by a trial witness, or a fresh witness coming forward
- 4) Fresh ‘real evidence’ – for instance documentary evidence
- 5) Other evidence⁷⁹²

Table 7.5 shows the number of times each category of fresh evidence was raised and the frequency that each category of fresh evidence was successful for murder and rape.

Table 7.5: Categories of fresh evidence

Category	Raised overall	Successful ground murder	Successful ground rape
1	42	11	6
2	23	13	2
3	36	1	9
4	25	7	10
5	4	0	0

Some differences become apparent in this data. Fresh expert evidence, in the form of either psychiatric evidence or other expert evidence, was most frequently successful in murder appeals. As discussed below, convictions for manslaughter were often substituted in these cases. 23 of the 55 successful murder appeals (41%) were successful on the basis of some kind of fresh expert evidence, whilst only 10% of successful rape appeals were for this reason. Successful fresh evidence in rape appeals were more likely to be based upon fresh witnesses or statements, or other kinds of real evidence.

⁷⁹² For instance, the fresh evidence of police impropriety seen in *R v Steele and others* [2006] EWCA Crim 195, or jury irregularity following investigation by the CCRC seen in *R v Thompson and others* [2010] EWCA Crim 1623.

Fresh evidence appeals in murder often raised similar issues. There were 6 cases where the evidence of Home Office pathologist Dr Michael Heath (which would be category 1) was the issue.⁷⁹³ There were 18 murder appeals where fresh evidence was submitted to attempt to demonstrate that the appellant had a viable defence of diminished responsibility (category 2). Of these appeals, 11 were successful, and in all but one case, where a retrial was ordered,⁷⁹⁴ the murder conviction was replaced by a manslaughter conviction. Once the Court is satisfied that the appellant did have a 'disease of the mind' and that his or her responsibility was diminished,⁷⁹⁵ so long as he did not hide his condition to attempt to secure an acquittal at trial,⁷⁹⁶ success usually follows. This again demonstrates that the 'unsafety test' is, to some degree, mediated by rules. Often the evidence of diminished responsibility will be so inconvertible that the Court will 'have no option'⁷⁹⁷ but to allow the appeal and replace the murder conviction with a manslaughter conviction. The appeal of Kenneth Erskine, convicted of seven counts of murder and one count of attempted murder, was described as straightforward; the evidence of diminished responsibility was overwhelming.⁷⁹⁸

The CCRC is a particularly potent source of fresh evidence. In the sample of murder and rape appeals studied in this thesis, there were 44 appeals following a CCRC referral, 33 of which raised fresh evidence. Of the 33 CCRC referrals raising fresh evidence, 21 were successful. There were therefore 97 non-CCRC appeals which raised fresh evidence, 38 of which were successful. Outside of CCRC referrals, therefore, fresh evidence was successful 39.1% of the times it was raised. Whilst this does suggest that CCRC referrals are somewhat responsible for the much higher overall success rate of fresh evidence observed in this thesis, the 39.1% success rate in non-CCRC appeals remains significantly

⁷⁹³ *R v Boreman and others* [2006] EWCA Crim 2265; *R v O'Leary (aka Dwyer)* [2006] EWCA Crim 3222; *R v L* [2007] EWCA Crim 1750; *R v Stanley* [2008] EWCA Crim 603; *R v Simmons* [2009] EWCA Crim 741; *R v Ahmed* [2010] EWCA Crim 2899.

⁷⁹⁴ *R v Inglis* [2010] EWCA Crim 2269.

⁷⁹⁵ Section 2 of the Homicide Act 1957. The definition of the defence was altered by the Coroners and Justice Act 2009, s 52.

⁷⁹⁶ See *R v Latus* [2006] EWCA Crim 3187.

⁷⁹⁷ See *R v Dass* [2009] EWCA Crim 1208 [41].

⁷⁹⁸ *R v Erskine*; *R v Williams* [2009] EWCA Crim 1425 [95].

higher than that seen in previous studies. This analysis of the role of fresh evidence is discussed further in Chapter 8.

7.2.3 The ‘Insufficient Evidence, Lurking Doubt, Generally Unsafe’ ground of appeal

Grounds of appeal coded as ‘lurking doubt’ were rare in both Roberts’s and Heaton’s studies.⁷⁹⁹ In the present study, a single ground of appeal was coded for any appeals which claimed that the verdict was unsafe due to there being insufficient evidence; appeals claiming the verdict was ‘generally unsafe’; and appeals which referred specifically to ‘lurking doubt’. Roberts and Heaton coded some of these kinds of grounds of appeal separately. While the Court does not often refer to the words ‘lurking doubt,’ both Roberts and Heaton note that there are other kinds of appeals which are equivalent to lurking doubt appeals. Heaton refers to there being a category of appeals which require the Court to consider the merits of the evidence: fresh evidence, issues of identification, and arguments of no case to answer, as well as lurking doubt.⁸⁰⁰ In Heaton’s study, there were 13 appeals based upon issues of identification in addition to cases coded as lurking doubt. Roberts coded appeals claiming ‘weak or insufficient evidence’ separately from ‘lurking doubt’ appeals, and found there were 8 of those appeals in her study, in addition to her lurking doubt appeals.⁸⁰¹

In the present study, there were 30 ‘generally unsafe, insufficient evidence, or lurking doubt’ appeals, meaning it was claimed in 6% of appeals, or 3% of grounds of appeal raised. Roberts’s ‘lurking doubt’ and ‘weak or insufficient evidence’ grounds represented 2% of her grounds; Heaton’s ‘lurking doubt’ and ‘weak ID evidence’ grounds represented close to 4% of his grounds. Thus, all three studies show that grounds of appeal most similar to ‘lurking doubt’ are rarely argued on appeal. Furthermore, such grounds are rarely successful. It was successful on just three occasions in the present study. It was successful only

⁷⁹⁹ Roberts (2009) 114, Heaton (2013) 123.

⁸⁰⁰ Heaton (2013), 124.

⁸⁰¹ Roberts (2009) 64.

once (for the 'weak ID evidence' ground) in Heaton's and three times (twice for weak evidence, once for lurking doubt) in Roberts's study. Such low numbers for success across all these studies show that the Court is indeed slow to quash convictions where it is necessary to go against the verdict of the jury in the absence of particular errors at trial.

The three cases where the appeal was allowed on the basis of this ground were two murder appeals and one rape appeal from 2010: *R v Haigh*,⁸⁰² *R v Azam and others*⁸⁰³ and *R v J*.⁸⁰⁴ In *Haigh*, Court held that the judge was correct to have allowed 'bad character' evidence in, and he was correct to have rejected a claim of no case to answer. The conviction for murder was quashed as there was insufficient evidence of murder rather than manslaughter following the death of the appellant's baby. In *Azam* convictions for joint enterprise murder gained after a third trial were quashed as the evidence relating to the identity of the shooter was unclear. Although the Court did not itself refer to 'lurking doubt' in its decision it did refer to the statement from the single judge Sir Christopher Holland granting leave explicitly on the basis of lurking doubt. In *J*, the conviction for rape was quashed, after finding (just) that there was a legitimate chain of reasoning meaning that the verdicts were not inconsistent, that 'this is one of those rare cases where on the particular facts and circumstances of the case the verdict on [the rape count] is unsafe.'⁸⁰⁵

As Heaton says, there are other grounds of appeal which essentially ask the Court to weigh up the evidence against the appellant in a similar way to lurking doubt appeals.⁸⁰⁶ The most important is fresh evidence which has already been discussed, but there is another important ground which operates in a similar way. Appeals which claim that the judge should have accepted a claim of no case to answer are similar to lurking doubt appeals and arise relatively frequently. The

⁸⁰² [2010] EWCA Crim 90.

⁸⁰³ [2010] EWCA Crim 226.

⁸⁰⁴ [2010] EWCA Crim 1768.

⁸⁰⁵ *ibid* [25].

⁸⁰⁶ Heaton (2013) 124.

ground was raised in 50 cases in the present study. Like lurking doubt appeals, it is also unlikely to be successful overall: just twice in the present study. It may be concluded therefore that there is relatively little difference as to how murder and rape appeals are treated based on this ground compared with appeals generally. As was discussed in Chapter 4, if the Court finds that the judge should have accepted a claim of no case to answer, it follows logically and as a matter of law that the quashing of the conviction must follow. Accordingly, it would appear that the Court is slow to find that this error occurred, because whilst it was raised relatively frequently, it was rarely successful. This suggests a relatively high degree of deference to the decision of the trial judge not to accept a claim of no case to answer.

The result of this element of the research suggests that when the Court is asked to weigh up the evidence against the appellant, in the absence of fresh evidence or procedural irregularities, they will rarely favour the appellant. Cases which could be grouped in an 'insufficient evidence' category are rarely successful. The Court of Appeal will usually defer to the judgment of the trial judge in no case to answer appeals, with the knowledge that doing otherwise means the conviction must be quashed. While this may be evidence of a restrictive approach, it may also make sense that convictions are rarely quashed on this basis. If the appellant was convicted of the offence by a jury, it may be expected that it will be rare for a Court of three judges, who did not hear all the evidence, to find there was in fact insufficient evidence to convict. As in previous studies, it is rare for the lurking doubt / insufficient evidence ground of appeal to be the only ground of appeal. It tends to be argued by counsel as a final attempt to rescue an otherwise failing appeal. These statistical findings will be discussed in Chapter 8. In particular it will be discussed whether the allegation of a restrictive approach is substantiated by these results.

This analysis feeds into the 'Did Error Occur?' legal variable which is utilised in this study. This substantiates that if no error occurred, and there is no fresh evidence, appeals are unlikely to be successful. Whilst previous studies have

suggested that this demonstrates a restrictive approach, it is not clear how this conclusion is reached. In Chapter 4, it was argued that the ‘unsafety test’ appears to provide the Court with fairly wide discretion, but the Court has interpreted the test so as to mean that if the Court thinks that the jury would still have convicted, had they known what the Court knows by the time of the appeal, they must allow the appeal. This, it was argued, demonstrates that the discretion offered by the unsafety test is framed by the necessity to operate the unsafety test in a way which upholds the purpose of the Court and the unsafety test – to correct miscarriages of justice and to uphold the rule of law. If it is shown that this is what occurs in this study of the Court, as measured by the ‘Did Error Occur?’ variable, this could be considered a measure of the legal model, and so impartiality, as the Court is operating based upon its understanding of the purpose of the ‘unsafety test’.

7.3 Summary of findings

The results of this element of the study broadly confirm the findings of previous studies, but there are some differences. Appeals based on procedural irregularities dominate grounds of appeal in murder and rape appeals, but not to the same extent as seen in previous studies. Fresh evidence has a far greater role in murder and rape appeals than in the general universe of appeals. This suggests a need for caution in drawing conclusions as to the general decision-making of the Court of Appeal given the demonstrable difference in approach to different crimes. In some ways the statistics for murder and rape paint a somewhat different picture of decision-making in the Court of Appeal, in particular in relation to fresh evidence. However, this study also shows that the Court will rarely quash convictions in appeals which require the Court to overrule the jury’s finding of fact. It is also rare for the Court to overrule a judge’s finding of sufficient evidence.

The remainder of this chapter explores the impartiality of the Court, and analyses the results of the relationship between a range of independent variables and the outcome of appeals against conviction.

7.4 Purposive selection procedure

There were over 100 individual pieces of data collected for each case including factual and demographic details and grounds of appeal. Some of the data were never intended to be used as variables in the binary logistic regression analysis but was intended for use as descriptive statistics or to aid in the creation of new variables or for identification purposes. Field discusses the rule of thumb that for binary logistic regression there should be approximately 10 events per variable (EPV) included in the analysis.⁸⁰⁷ Peduzzi et al showed that having fewer than 10 EPV can lead to biased or overfitting results due to there being too many variables compared to the number of events.⁸⁰⁸ It is the number of successful appeals which is considered when determining the number of 'events'.⁸⁰⁹ This is because it is the outcome which is least likely to occur which should be considered when calculating EVP. This means the relevant number of 'events' for the purposes of EPV is the 135 successful appeals. Thus, a maximum of 13 independent variables, if following the 10 EPV rule of thumb, could be included in the final analysis to reduce the chance of biased results. As Field explains, the 10 EVP rule of thumb oversimplifies the issue of variable selection.⁸¹⁰ For instance, Vittinghoff and McCulloch found that as low as 5-9 EPV could be cautiously used.⁸¹¹

As was shown in Chapter 6, there are more than 13 independent variables in this study. As a result, some of the variables cannot be utilised in the final modelling. Usually, this selection exercise is not necessary because either the dataset is very large, or there were a large number of events, or there were fewer or more specific variables collected. For instance, if there were two variables and 200 'events', no selection process would be necessary. Alternatively, if the

⁸⁰⁷ A Field, *Discovering Statistics Using IBM SPSS Statistics* (4th ed, SAGE Publishing 2013) 313.

⁸⁰⁸ P Peduzzi, J Concato, E Kemper, TR Holford and AR. Feinstein 'A Simulation Study of the Number of Events Per Variable in Logistic Regression Analysis' (1997) 49 (12) *J Clin Epidemiology* 1373, see also MA Babyak 'What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models' (2004) (66) *Psychosomatic Medicine* 411.

⁸⁰⁹ *ibid.*

⁸¹⁰ Field (n 807) 313.

⁸¹¹ E Vittinghoff and CE McCulloch, 'Relaxing the Rule of Ten Events Per Variable in Logistic and Cox Regression' (2007) (165) *Am J Epidemiology* 710, 717.

researcher had access to a large pool of research participants, and a pre-defined number of variables, it would be possible to determine how large a sample was needed based on the expected power of the variables. For instance, a researcher could determine the size of the sample needed if it was intended to collect ten variables from each case, based on an estimation of power. As Field shows, if a researcher intends to collect ten variables and she expects a medium effect, she would need to collect a sample of approximately 119 cases.⁸¹²

As there have been no similar studies of this nature conducted on the Court of Appeal, except for the 'Did Error Occur?' variable, it was not possible to know which variables were likely to be associated with outcomes and which were not. Accordingly, the purpose of this research is to address that question as an initial explorative investigation. Later researchers can utilise the findings of this thesis to conduct further research on the Court of Appeal or other courts.

The method used to select variables for further analysis is the 'purposive selection' procedure suggested by Hosmer, Lemeshow and Sturdivant.⁸¹³ They suggest conducting univariate analysis on each independent variable separately. Any variable with a p-value of 0.25 or higher would be omitted from the analysis. Since, as was explained above, the p-value is a relatively weak analytic test of the truth of a null hypothesis, this higher p-value cut off means any variables dismissed at this stage are likely to have little predictive power. However, as explained in Chapter 5, for variables with larger p-values this should not be interpreted as a finding of no association, or an acceptance of the null hypothesis. Similarly, variables with p-values smaller than 0.25 should not be interpreted as having substantive effects on outcomes. Rather, the p-value should be considered a descriptive finding of the strength of correlation rather than inferential.

⁸¹² Field (n 807) 314.

⁸¹³ DW Hosmer, S Lemeshow and RX. Sturdivant *Applied Logistic Regression* (3rd ed, Wiley 2013) 89-94.

The difficulty with the purposive selection method is that some variables may be insignificant on their own but in combination with others may have a stronger association with outcomes.⁸¹⁴ The alternatives suggested by Babyak are to collect more data or to combine variables. In this study it was decided that the sample size was large enough to not require further collection of data, and there were no variables to combine. As Hosmer and his colleagues state, by using a higher p-value than the more conventional 0.05 the risk of omitting potentially relevant variables is mitigated.⁸¹⁵ An analysis by Bursac and colleagues suggested that this was a useful and viable technique for covariate selection.⁸¹⁶ This method was useful in removing some variables which had very large p-values (i.e. non-significant) which would be less likely to contribute to the final model as an attempt to model outcomes. As this is an exploratory study, this initial univariate screening exercise was helpful at identifying those variables which might, and might not, have an association with the judges' decision-making.

Table 7.6 shows the results of the purposive selection procedure. This shows the p-value for each variable and the co-efficient (*B*) for each variable. If the co-efficient has a negative sign this indicates that the variable is associated with having an appeal dismissed; a positive sign indicates that the variable is associated with a successful appeal. The further the co-efficient is from zero, in either direction, the stronger predictor the variable is likely to be. Full statistics for each variable included in this stage is provided in Appendix C. The 'Did Error Occur?' variable was not included in the purposive selection procedure because it has already been hypothesised that it would have a strong relationship with outcomes, and it had been determined previously that the variable would be included in the analysis.

⁸¹⁴ See MA Babyak 'What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models' (2004) (66) *Psychosomatic Medicine* 411, 417.

⁸¹⁵ Hosmer, Lemeshow and Sturdivant (n 813) 91.

⁸¹⁶ Z Bursac, CH Gauss, DK Williams and DW Hosmer, 'Purposeful Selection of Variables in Logistic Regression; (2008) (3) *Source Code of Biology and Medicine* 17.

Table 7.6: Purposive selection procedure results. Cut off: $p = > 0.25$.

Variable Name	<i>P value</i>	<i>B</i>
<i>1) Joined case (*) (n = 55)</i>	<i>0.13</i>	<i>-0.52</i>
Multi Murder / Rape (n = 116)	0.505	0.155
Convicted Other Offences (n = 195)	0.714	-0.76
Monday (n = 39)	0.49	0.24
Tuesday (n = 106)	0.57	-0.14
<i>2) Wednesday (*) (n = 108)</i>	<i>0.09</i>	<i>-0.42</i>
Thursday (n = 106)	0.68	0.09
<i>3) Friday (*) (n = 113)</i>	<i>0.17</i>	<i>0.31</i>
January (n = 39)	0.49	0.24
<i>4) February (*) (n = 46)</i>	<i>0.18</i>	<i>0.42</i>
March (n = 46)	0.69	-0.14
<i>5) April (*) (n = 29)</i>	<i>0.17</i>	<i>-0.69</i>
May (n = 59)	0.72	0.10
June (n = 44)	0.57	-0.20
July (n = 64)	0.92	-0.02
August (n = 7)	0.41	-0.88
September (n = 11)	0.44	-0.60
October (n = 37)	0.87	0.05
November (n = 43)	0.34	0.32
December (n = 47)	0.62	-0.17
Extempore (n = 307)	0.79	0.05
QC Appellant Only (n = 54)	0.85	0.05
<i>6) QC Crown Only (*) (n = 35)</i>	<i>0.24</i>	<i>-0.50</i>
F Counsel Appellant (n = 99)	0.86	0.04
F Counsel Crown (n = 119)	0.35	0.21
F Judge Present (n = 133)	0.82	0.04
LCJ Present (n = 29)	0.76	0.12
VP Present (n = 39)	0.67	-0.16
President Present (n = 24)	0.39	-0.44
Circuit Judge Present (n = 95)	0.42	-0.21
Recorder Present (n = 60)	0.50	-0.20
<i>7) Retired Judge Present (*) (n = 52)</i>	<i>0.18</i>	<i>0.41</i>
Trial Judge Sat in CA (n = 101)	0.13 but see in discussion Chapter 6	-0.404

Variable Name	<i>P value</i>	<i>B</i>
Judge 1 (n = 25)	0.33	-0.49
<i>8) Judge 2 (*) (n = 20) Becomes Judge 1</i>	<i>0.03</i>	<i>0.96</i>
<i>9) Judge 3 (*) (n = 32) Becomes Judge 2</i>	<i>0.04</i>	<i>-1.08</i>
<i>10) Judge 4 (*) (n = 43) Becomes Judge 3</i>	<i>0.00</i>	<i>0.97</i>
Judge 5 (n = 34)	0.77	-0.11
Judge 6 (n = 39)	0.75	0.11
Judge 7 (n = 20)	0.88	0.07
Judge 8 (n = 29)	0.47	0.29
Judge 9 (n = 23)	0.78	-0.13
Judge 10 (n = 25)	0.60	-0.25
Judge 11 (n = 26)	0.28	-0.54
White Brit (n = 83)	0.04 see relevant note in Chapter 6	0.66
Cases cited	0.81	-0.00
<i>11) Offence Rape (*) (n = 231)</i>	<i>0.00</i>	<i>-0.58</i>
<i>12) Low Sentence (*) (n = 76)</i>	<i>0.09</i>	<i>0.45</i>
Medium Sentence (n = 213)	0.93	-0.01
High Sentence (n = 86)	0.95	0.01
<i>13) Very High Sentence (*) (n = 48)</i>	<i>0.05 but see relevant note in Chapter 6</i>	<i>-0.83</i>
Appellant Good Character Stated (n = 88)	0.46	0.18
<i>14) Appellant Bad Character Stated (*) (n = 210)</i>	<i>0.00</i>	<i>-0.65</i>
Denial of AR (n = 284)	0.50	-0.13
<i>15) Denial of MR (*) (n = 165)</i>	<i>0.14</i>	<i>0.30</i>
<i>16) CCRC (*) (n = 44)</i>	<i>0.00</i>	<i>1.128</i>
<i>17) Granted Single Judge (*) (n = 278)</i>	<i>0.00</i>	<i>-0.65</i>
<i>18) Historical Offence (*) (n = 57)</i>	<i>0.07</i>	<i>0.51</i>
Appellant under 18 (n = 49)	0.98	-0.00
<i>19) Age Under 13 (*) (n = 108)</i>	<i>0.08</i>	<i>0.402</i>
Age 13 – Under 16 (n = 33)	0.53	0.239
Age 16 - 17 (n = 34)	0.91	0.04
<i>20) Age 18+ (*) (n = 296)</i>	<i>0.04</i>	<i>-0.42</i>
Drink / Drugs (n = 161)	0.27	-0.237

Variable Name	<i>P value</i>	<i>B</i>
Victim Male (n = 190)	0.62	-0.10
<i>21) Unanimous Verdict (*) (n = 390)</i>	<i>0.02 but see relevant note in Chapter 6</i>	<i>-0.54</i>
Known Victim (n = 393)	0.87	0.04

Fuller analysis of this Table is deferred until Chapter 8, but there are some significant points to note from this initial analysis. Some variables of note had large p-values, suggesting weak association. For instance, the Female Judge variable had a p-value of 0.82. As was discussed in Chapter 3, gender has been frequently used as a variable in ELS studies. This result therefore adds to the already mixed picture of the association between of gender and judging. There were 70 murder appeals in which there was at least one female judge, and 17 (24%) of these were successful. There were 80 rape appeals in which there was at least one female judge, and 22 (27%) of these were successful. There is little difference in the success rates between the two offences when a female judge is present, therefore suggesting little association. If gender could be considered to be an issue in rape appeals, it does not appear to have any bearing on the outcome of appeals against conviction with female judges.

Other variables of note which had larger p-values were the ranks of the judges in the judicial hierarchy. Only the Retired Judge variable made it through the purposive selection procedure. Thus, an appellant, or the Crown, has little to fear if they get more or less senior judges; there does not appear to be a strong association. In Darbyshire's qualitative observations of the Court she found that judges of lower ranks did not simply defer to higher ranks, and could observe no patterns in results.⁸¹⁷ The present study supports this finding.

It is also interesting to note some of the signs of the coefficients, even for variables which were not strong predictors. For instance, if the trial judge had

⁸¹⁷ P Darbyshire, *Sitting in Judgment: The Working Lives of Judges* (Hart Publishing 2011) Chapter 14.

previously sat in the Court of Appeal, this was associated with an unsuccessful appeal. As is noted in Chapter 6, however, this variable was not included in the later models due to missing data. It is also pertinent to note the opposite signs of very low and very high sentences. Very low sentences were associated with success and very high sentences were associated with failure. It will also be noted that some judges were associated with successful appeals whilst some were associated with unsuccessful appeals.

The majority of the independent variables analysed in this study are shown to be only weakly associated with the outcomes of appeals against conviction. As was stated in Chapters 5 and 6, each variable had a null hypothesis that the variable is not associated with outcomes. It was also explained that for most variables it is unlikely to ever be the case that there is no association at all with the dependent variable. This was shown to be the case, as no variable had a B of zero, after noting that the B is rounded to three decimal places. Accordingly, it would be inaccurate to state that these variables had no association at all with the outcome of appeals. However, in this study, it has been sought to identify the variables which are more strongly associated with outcomes. Only variables with a p-value of lower than 0.25 were thus included in the next stage of analysis.

7.5 Results of initial multiple binary logistic regression analysis

As an initial assessment, all the 21 variables with p-values smaller than 0.25 were analysed in a binary logistic regression model. This caused a reduction in the number of cases included in the analysis from 472 cases, to 423 cases. The cause of this reduction is the sentencing variables. The sentencing variables were subject to missing data as not all judgments would state clearly what the sentence was. In the 423 cases the split between outcomes was 311 (73.5%) appeals dismissed and 112 appeals allowed. This means that the missing data for sentencing appears to disproportionately impact successful appeals, as the percentage of successful appeals in the smaller dataset is lower compared with the full dataset. This is a problem for two reasons: successful appeals were already relatively rare which makes them more difficult to predict; and the lower

dataset size may affect the EPV. To reduce the chance of biased results, 11-13 variables could be included in this dataset. How this was dealt with is discussed below.

Despite this difficulty, the results of the initial binary logistic regression model are shown below. The overall significance of the model was less than 0.000, meaning that the model performs significantly better than the constant only model. A Hosmer and Lemeshow goodness of fit test was conducted. This tests the null hypothesis that the model is a good fit for the data. The p-values of 0.727 meaning that the model is a good fit for the data. The Cox and Snell and Nagelkerke pseudo R² were 0.14 and 0.21 respectively. This means that the model explains between 14% and 21% of the variation in outcomes. This shows that the 21 variable model performs moderately well at predicting outcomes. This means that this model would be unlikely to be of use in predicting outcomes in particular cases, but it does allow an assessment of the strength of the association between the variables and the outcomes.

The following Table shows the number of successful and unsuccessful appeals which were correctly classified by this model.

Table 7.7: Classification table: Initial analysis.

	Appeals Allowed	Appeals Dismissed	
Correctly Classified	30 (26.8%)	294 (94.5%)	
Overall Correct			324 (76.6%)

It will be seen that the model performs very well at predicting dismissed appeals but only moderately at predicting successful appeals. However, of the 423 cases included in this model, some 375 were predicted to have been dismissed based by the model. This may explain why model is able to correctly classify a very high proportion of unsuccessful appeals. Since most appeals were predicted to have been dismissed, and most appeals were dismissed, by chance it will correctly

predict a large amount of unsuccessful appeals. It would appear therefore that this model does not perform especially well at discriminating between successful and unsuccessful appeals, and the variables are poor at predicting an allowed appeal. This would suggest that, while overall the model can correctly classify 76.4% of appeals, the model is not able to differentiate between successful and unsuccessful appeals to any great extent.

Table 7.8, below, shows the variables which were entered into the model. The *B* column is the co-efficient for each variable. The S.E column is the standard error of the co-efficient. In this model the standard errors are relatively small and robust, i.e. none are approaching 1. The Sig. column is the p-value for the individual variables. The Exb(b) column is the odds ratio (OR), which explains the change in the likelihood of a successful appeal if that variable is present. An odds ratio below 1 indicates that if the variable is present the odds of a successful appeal decrease by the value of the odds ratio; an odds ratio of above 1 indicates that if the variable is present the odds of success increase by the value of the odds ratio. Variables marked with an asterisk in the Sig. column indicate that the variable is a statistically significant predictor at the 5% level.

Table 7.8: Variables in equation: Initial analysis.

Variable	B	S.E	Sig.	Exp(B)
Joined Case	-0.265	0.445	0.551	0.767
Wed	-0.223	0.318	0.483	0.800
Fri	0.288	0.309	0.351	1.334
Feb	-0.147	0.444	0.741	0.864
Apr	-0.803	0.570	0.159	0.448
QC Crown Only	-0.953	0.529	0.072	0.386
Retired Judge Present	0.141	0.374	0.706	1.152
Judge1	1.230	0.560	0.028*	3.421
Judge2	-1.164	0.659	0.077	0.312
Judge3	0.959	0.417	0.021*	2.610
Offence Rape	-0.194	0.326	0.552	0.824
Sentence Low	0.600	0.311	0.054	1.822
Sentence V High	-0.317	0.477	0.507	0.729
Bad Character	-0.367	0.273	0.179	0.693
Denial of MR	0.818	0.289	0.005*	2.266
CCRC	1.167	0.459	0.011*	3.212
Single Judge	-0.660	0.265	0.013*	0.517
Historical Offence	0.863	0.388	0.026*	2.370
Under_13	0.304	0.423	0.472	1.356
Age_18+	0.018	0.378	0.962	1.018
Unanimous	-0.242	0.297	0.414	0.785
Constant	-0.913	0.505	0.070	0.401

* = significant at 5% level

This Table is shown largely for completeness and openness of data and the analytic process. There is no intention to further analyse this Table, because there are likely to be too many variables included in this model to mean the results are reliable. As such, confidence intervals are not provided at this stage.

In order to have the EPV at an acceptable level, several variables need to be removed. Only variables with a p-value below 0.500 are retained in the analysis initially, the rest are removed. This higher p-value was used in order to retain more potential variables in the model. This resulted in the removal of the following variables with p-values higher than 0.500:

- Joined case
- February

- Retired Present
- Rape
- Sentence very high
- Adult Victim

Note that this does not remove any cases from the sample, but removes that variable from the analysis. The remaining day of the week and months of the year variables were also removed from the model at this stage. This was because the majority of these variables were removed at the univariate stage, and they had only moderate association with outcomes in the above binary logistic regression analysis, and so it was concluded that these associations were more likely to be spurious. After removing the three remaining date variables, the number of variables is reduced to 12. Finally, it was decided to remove the 'low sentence' variable. This is because of the effect it has on the size of the dataset. It was also not statistically significant. Therefore, at this stage the model was returned to the full size of 472 appeals, with 11 variables. As there were 135 'events', the EPV was now over 10, thereby helping to ensure the robustness of the data analysis being undertaken.

7.6 Binary logistic regression Model 1

The 11 remaining variables were included in a further binary logistic regression model, the results of which are now presented. This will be called MODEL 1 throughout this chapter. Model 1 was statistically significant, with a p-value of less than 0.000. The Hosmer and Lemeshow goodness of fit test showed the model to be a good fit for the data, with a p-value of 0.852. The Cox and Snell and Nagelkerke pseudo R² results were 0.11 and 0.16 respectively. The classification Table for this model is shown below.

Table 7.9: Classification table: Model 1.

	Appeals Allowed	Appeals Dismissed	
Correctly Classified	33 (24.4%)	318 (94.4%)	
Overall Correct			74.4%

The variables included in this model and their p-values are shown in the Table below. Variables significant at the 5% level are marked with an asterisk.

Table 7.10: Variables in equation: Model 1.

Variable	B	S.E	Sig.	Exp(B)
QC Crown Only	0-.745	0.476	0.118	0.475
Judge1	1.025	0.504	0.042*	2.787
Judge2	-0.785	0.558	0.160	0.456
Judge3	1.034	0.345	0.003*	2.813
Bad Character	-0.531	0.233	0.023*	0.588
Denial of MR	0.532	0.245	0.030*	1.703
Historical Offence	0.730	0.345	0.034*	2.076
Under 13	0.215	0.292	0.461	1.240
CCRC	0.884	0.370	0.017*	2.420
Single Judge	-0.515	0.239	0.031*	0.598
Unanimous	-0.483	0.261	0.064	0.617
Constant	-0.540	0.337	0.109	0.583

* = significant at 5% level

The following Table shows the 95% confidence intervals for the exp(B) (odds ratio) of each variable. The 95% confidence interval can be interpreted as meaning that the odds of a successful appeal may change by as little as the lower bound of the interval, or as much as the upper bound, with 95% confidence.⁸¹⁸ For instance, the first entry in Table 7.11, QC Crown Only, means the model calculates with 95% confidence that the odds of a successful appeal may decline to 0.187 or increase to 1.207 if only the Crown was represented by a QC. It is possible that the true odds ratio is outside this range, but it is relatively unlikely if a 95% confidence interval is used.⁸¹⁹ A smaller range means that the variable is more robust, and more confidence can be had in the precision of the results. Variables in which the confidence interval crosses the boundary of 1, such as the QC Crown Only variable, will not be statistically significant, because an OR of 1 means there is no observed difference in outcome when that variable is present.

⁸¹⁸ See DW Hosmer, S Lemeshow and RX Sturdivant, *Applied Logistic Regression* (3rd ed, Wiley 2013) 54.

⁸¹⁹ There is disagreement and some controversy over the proper interpretation of confidence intervals. This interpretation is the most widely used, see G Cumming, *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis* (Routledge 2012) 80.

Accordingly, since the confidence interval operates at 95% confidence, and not 100% confidence, it is not possible to exclude the possibility that the odds ratio is actually 1, and so the variable will not be statistically significant. The further away the range is from 1, in either direction, the stronger the association.

Table 7.11 95% confidence interval for exp(B): Model 1.

Variable	Lower bound	Upper bound
QC Crown Only	0.187	1.207
Judge 1	1.038	7.479
Judge 2	0.153	1.363
Judge 3	1.431	5.530
Bad Character	0.373	0.929
Denial of MR	1.054	2.751
Historical Offence	1.055	4.083
Under 13	0.700	2.197
CCRC	1.171	5.000
Single Judge	0.374	0.954
Unanimous	0.370	1.029

As this is not the final model of the study, it is unnecessary to scrutinise these statistics in great detail, however, a number of points are important. Recall that the odds ratio is the change in odds of a successful appeal if that variable is present. Only one variable, at the lower bound, is very much above 1 – Judge 3. The other statistically significant variables, such as Judge 1, denial of MR, and CCRC are only just above 1 at the lower bound – meaning that it is possible their association with outcomes is minimal. On the other hand, the upper bound of some variables is much higher than 1, suggesting it is possible that the association is strong. However, as Hosmer, Lemeshow, and Sturdivant note, it is common for binary logistic regression confidence intervals to be right-skewed (i.e. high upper bounds) because the confidence interval can range from 0 to infinity.⁸²⁰ The confidence intervals in this model are frequently fairly large, suggesting that the model is imprecise.

⁸²⁰ Hosmer, Lemeshow, and Sturdivant (n 818) 54.

The pseudo R^2 statistics show that this model accounts for approximately 11% to 16% of the variation in the sample. This is within the range of R^2 statistics reported in other ELS studies of judicial decision-making. For instance, Sisk and Heise reported a score for Nagelkerke R^2 beginning at 18%, although at times rising to 40%.⁸²¹ Sisk, Heise and Morriss's logistic regression tests reported a McFadden pseudo R^2 (not reported in this thesis) of between 6% and 30%.⁸²² In Chew and Kelley's study of the influence of race on decision-making they reported a Nagelkerke R^2 of between 3% and 17%.⁸²³ Juliano and Schwab's study of sexual harassment cases reported a pseudo R^2 (it did not stipulate which test is reported) of between 11% and 20%.⁸²⁴ The most comprehensive use of R^2 as an assessment of the power of models is Cross's,⁸²⁵ he rarely received R^2 of above 15% for his numerous studies of variables. What this means is that the model could not be used to predict particular outcomes, particularly given the low level of correct classification if the appeal was allowed. The model is therefore best used to demonstrate that there is an association between particular variables and outcomes.

The binary logistic regression model shows that a number of the variables analysed are statistically significant at the 5% level in their measured effect on the prediction of Court of Appeal decisions. The statistically significant variables are: Judge 3; Bad Character; Denial of Mens Rea; Historical Offence; CCRC Reference; and Single Judge. The remaining predictors were not significant at the 5% level.

Using the odds ratios in Table 7.10 it can be seen that in an appeal with Judge 3 on the bench, the odds of having a conviction quashed increase by 2.666, and

⁸²¹ GC Sisk and M Heise 'Judges and Ideology: Public and Academic Debates About Statistical Measures' (2005) (99) *NWU L Rev* 743, at footnotes 266-7.

⁸²² GC Sisk, M Heise and AP Morriss, 'Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning' (1998) (73) *NYU L Rev* 1377, at 1433 - 48.

⁸²³ PK Chew and RE Kelley 'Myth of the Color-Blind Judge: An Empirical Analysis of Racial Harassment Cases' (2009) (86) *Wash U L Rev* 1117, footnote 143.

⁸²⁴ A Juliano and SJ Schwab 'The Sweep of Sexual Harassment Cases' (2001) (85) *Cornell Law Review* 548, 572.

⁸²⁵ FB Cross, *Decision Making in the US Courts of Appeals* (Stanford University Press 2007).

this is statistically significant. Although not statistically significant, if Judge 1 was on the bench the odds increase by 2.826. The odds of having a conviction quashed increase by 2.459 if the case was a CCRC reference; but decline to 0.594 if leave was granted by the single judge. Interestingly, the odds of having a conviction quashed are increased by 1.625 if the defence at trial was a denial of *mens rea*, while denying the *actus reus* was excluded at the univariate stage. Intuitively this seems to make sense from a legal perspective. It is much more difficult to prove a person's state of mind in at the time of allegedly committing a criminal offence. It follows that the Court of Appeal may be more inclined to quash convictions where the *mens rea* is in dispute because definitive proof of a mental state is very difficult to obtain.

As was discussed above, however, the confidence intervals for the odds ratios are for some variables fairly wide. Some variables range from having potentially very little association with outcomes to potentially a strong association with outcomes. As such, at this stage it appears that the model lacks precision.

It is useful to look at the variables and consider whether the outcome of the binary logistic regression analysis has face validity; that is, whether it makes intuitive sense. In particular, it is useful to look at the exp(B) column and ask whether the direction of the change in odds is what would be expected. All of the variables included in the above model make intuitive sense. It makes sense that having a QC for the Crown and not for the appellant could lead to lower odds of having a conviction quashed; it would require a strained interpretation for increased odds to make intuitive sense. It also makes sense that the inclusion of some judges on a bench leads to increased odds, and the inclusion of others (i.e. Judge 2) leads to reduced odds. It makes sense that if the judgment states that the appellant has bad character the odds of having a conviction quashed are reduced. Historical offences are shown to be associated with increased odds of having convictions quashed. This may be understandable, convictions relating to offences committed many years in the past are more suspect; it follows that the Court of Appeal may be more inclined to quash them. CCRC referrals have

already gone through a significant review process and so it follows that they may be more likely to be quashed.

If leave was granted by the single judge, as opposed to being a CCRC referral, or leave having been granted by the full court or by the trial judge (which is very rare), the odds of having a conviction quashed are reduced. Again, this arguably makes intuitive sense. The majority of appeals are granted leave by the single judge and the majority of appeals are dismissed. It may suggest that the single judge is too ready to give leave, or the Court is generally conservative in quashing convictions without closer scrutiny by the Full Court or the CCRC. Finally, if the verdict was unanimous the odds decrease. This makes sense: if the jury was unanimous it follows that the Court may be less inclined to allow the appeal.

As has been discussed previously in this thesis, some variables may be indicators of a lack of impartiality because it would be improper for them to be associated with outcomes. The judges on the bench are shown to be predictors of success or failure; this could be considered a behavioural variable, and may be considered contrary to the legal model of judicial decision-making. Alternatively, this underlying relationship between judges and outcomes may relate to the institutional factors which determine how benches of judges are composed. The previous bad character of an appellant should not be associated with outcomes, but this model shows that bad character is associated with having lower odds of a conviction being quashed. As has been stated previously, this is insufficient as evidence to conclude that the Court lacked impartiality in its decision-making. The aim of this study was as an initial exploration of patterns in the data, and in this firm model there are indeed patterns in the data.

As was discussed in Chapter 6, the CCRC referral variable is included in the study as a counterweight to factual and demographic variables. It is known that CCRC appeals are more likely to be successful, and so it may appear reasonable that CCRC referrals are statistically significant. CCRC referrals have previously

been through a lengthy review process, and the CCRC will have concluded that there is a real possibility that the conviction will be found unsafe. As this variable is found to be statistically significant, and has some relationship with the odds of successful appeals, it would appear that the institutional relationship between the CCRC and the Court of Appeal is observable in the Court of Appeal data.

Table 7.7 shows that this model can correctly predict 94.1% of cases which were dismissed but only 21.5% of cases which were allowed. As Hosmer, Lemeshow and Sturdivant note,⁸²⁶ it is quite common that binary logistic regression models perform better in predicting one outcome than another. This is particularly likely to occur when one outcome occurs considerably less frequently than another.⁸²⁷ The 94.1% correct figure is likely to be artificially high due to the relatively high number of appeals which were dismissed overall. The more interesting finding is that the model can predict only 21.5% of successful appeals. This represents only a moderate ability to correctly predict successful appeals. This suggests that there may be an alternative variable or factors which helps to explain successful appeals. One additional variable which could account for success is the law which governs the case.

7.7.1 The law governing the case

Model 1 explored the relationship between a range of variables on the question of whether the conviction is unsafe, and it has been shown that some variables are associated with particular outcomes. It has been explained previously in this thesis that a variable designed to capture the law has been created. The measure used is the 'Did Error Occur?' variable. In the data collection exercise, the 'Did Error Occur?' variable was answered 'Yes' if the Court decided that an error occurred which required them to decide whether the error made the conviction unsafe. Accordingly, the variable was answered 'No' if the Court found

⁸²⁶ DW Hosmer, S Lemeshow and RX Sturdivant *Applied Logistic Regression* (3rd ed, Wiley 2013) 170-1.

⁸²⁷ *ibid.*

that there was no error in the proceedings, or that the error was so minor that it was not necessary to consider whether it made the conviction unsafe.

The following Table shows the relationship between the 'Did Error Occur?' variable and the 'Conviction Quashed' variable.

Table 7.12: Number of appeals allowed / dismissed, by whether an error occurred.

	Error Occurred	Error Did Not Occur
Appeals Allowed	86	49
Appeals Dismissed	59	278
Total	145	327

As Table 7.12 shows, there were 145 cases where an error occurred, and 327 appeals where the Court found that no error occurred. Of these 327, 278 appeals were subsequently dismissed. It will be observed that this is by far the most likely occurrence. This suggests that a decision that an error did not occur has a significant association with the outcome of appeals. It further suggests that when an error did not occur appeals are very likely to be dismissed. When the variable 'Did Error Occur?' was answered 'Yes' there were 59 appeals dismissed. When the 'Did Error Occur?' variable was answered 'No' there were 49 appeals allowed. All but one of these latter appeals was a fresh evidence appeal. When the variable 'Did Error Occur?' was answered 'Yes' there were 86 appeals allowed. The significance of this is considered below.

The 'Did Error Occur?' variable may provide some measure of law on decision-making. It is bound to be incomplete because it is only one facet of 'law' but it may provide some insight. It is considered to be a legal variable because the question is usually addressed by reference to legal rules. It is a legal variable because the Court will determine what the existing law is and determine whether the law was correctly applied in the facts of the case.

In order to evaluate the legal variable, a further binary logistic regression analysis was conducted to explore whether the variables which passed the initial 'purposive selection' screening were predictors of the answer to the question: 'Did an error occur'? The same 21 independent variables as were included in the initial multivariate model were included in this model, this time the outcome variable was the 'Did Error Occur?' variable. The justification for conducting a further binary logistic regression model with an amended outcome variable is that the decision-making of the Court of Appeal is a distinctly two-stage process. In most appeals the Court will have to decide whether an error occurred, and then decide whether the conviction is unsafe. This model is testing whether the variables are predictors of the answer to this first question.

This model performed poorly at predicting outcomes and few of the variables were statistically significant. Because this model had the sentencing variables the dataset size was again reduced, the full results are not reported here. Overall correct classification fell to 70.2%, with only 8% of cases where an error occurred correctly classified. The Cox and Snell, and Nagelkerke R^2 statistics were 0.063 and 0.09 respectively. Only one variable was a significant predictor at the 5% level: bad character.

To complete this analysis of the legal variable, a final binary logistic regression analysis was conducted on the 'Did Error Occur?' variable using the same 11 variables as were included in Model 1. This will be called MODEL 2 in the remainder of this chapter. Model 2, now with the full dataset of 472 cases, showed no improvement on the previous model in terms of goodness of fit and accuracy. The pseudo R^2 range was 0.04. and 0.06, meaning 4% to 6% of the variation in the outcome was explained by the model. The classification Table for this model is below:

Table 7.13: Classification table: Model 2.

	Error Did Occur	Did Not Occur	
Correctly Classified	12 (8.3%)	312 (95.4%)	
Overall Correct			68.6%

It will be observed that the correct classification of cases which were found to contain errors was 8.3%, with 68.6% of decisions correctly classified overall. There were only 12 appeals correctly classified by the model as being answered 'Yes', suggesting that the variables are poor predictors of whether the Court would find that an error occurred.

The significance of the variables is shown below. Variables significant at the 5% level are marked with an asterisk. Owing to the limited power of this model overall, confidence intervals are not considered necessary.

Table 7.14: Variables in equation: Model 2.

	B	S.E	Sig.	Exp(B)
QC Crown Only	-0.636	0.449	0.157	0.530
Judge 1	0.395	0.495	0.425	1.484
Judge 2	-0.341	0.450	0.448	0.711
Judge 3	0.884	0.334	0.008*	2.420
Bad Character	0.205	0.215	0.341	1.227
Denial of MR	0.295	0.229	0.198	1.343
Historical Offence	0.883	0.331	0.008*	2.418
Under 13	-0.326	0.294	0.267	0.722
CCRC	-0.354	0.402	0.379	0.702
Single Judge	-0.166	0.224	0.460	0.847
Unanimous	-0.368	0.251	0.143	0.692
Constant	-0.680	0.323	0.035	0.506

* = Significant at 5% level

7.7.2 Discussion

It will be observed that two of the 11 variables remain significant at the 5% level: Judge 3, and Historical Offence. Both have positive coefficients (*B* column) meaning that these two variables are predictors of the Court finding that an error did occur. This may suggest that historical offence trials were more likely to be found to have had procedural irregularities. This is a noteworthy point, because a common argument in historical offences was that the trial should never have occurred due to the delay in bringing the proceedings being an abuse of process.

Some of the variables change from being significant in Model 1, to being non-significant in Model 2. For these variables, there was a statistically significant relationship with the question of unsafety, but not with the question of whether an error occurred. The overall statistics for Model 2 show that the variables are weak predictors of the answer to the question 'Did Error Occur?' This data is potentially revealing regarding decision-making in the Court of Appeal. The independent variables show limited ability to predict the outcome of the answer to the first question, 'Did the error occur?' Conversely, the independent variables showed greater predictive power in relation to the second question 'Is the conviction unsafe?' It must be recalled, however, that the power of Model 1 was still fairly weak. This may suggest that the question of whether an error occurred has less of an association with the independent variables. This may suggest that the question of whether an error occurred is answered in a more objective, or impartial, manner.

Although one judge was shown to be a statistically significant predictor, the overall predictive power of Model 2 does not suggest that there is any substantive significance to that judge. This may be because the question 'Did the error occur?' is more clearly a legal question than the overall question of unsafety. Statutes and common law lay down rules and the Court of Appeal must determine whether they were correctly applied by the Court below. This decision is usually made by reference to the statutes or precedent and it appears that this question will usually be the same regardless of which variables are present.

This suggests that the answer to the question ‘Did the error occur?’ appears not to be associated strongly with factual and demographic variables. It also confirms that when the ‘Did Error Occur?’ variable is answered yes, the appeal will often be successful. This may partly explain why Model 1 performs only moderately well at predicting allowed appeals: the majority of allowed appeals are settled by the answer to the question ‘Did the Error Occur?’ This supports the statistical conclusions of previous studies, in relation to the presence of an error being a strong predictor of the outcome of appeals against conviction. As has been explained previously, however, it is not argued that this is strong evidence of the ‘restrictive approach’ of the Court, but could be interpreted as the Court following the law.

As Table 7.12, above, shows, in most appeals the answer to ‘Did the Error Occur?’ is ‘No’, and usually this will lead to the appeal being dismissed. This confirms Roberts’s finding that a finding of no error usually determines appeals.⁸²⁸ But as is also shown by Table 7.12, there is a split between appeals which were successful when an error occurred, and appeals which were unsuccessful when an error occurred. There were also appeals which were successful when there was no error found at trial.

This is partly explained by the fact that all but one of the 49 cases which were allowed when there was no error at trial were fresh evidence appeals. As was discussed in Chapter 3, the power to admit and accept fresh evidence is discretionary and requires careful judgment by the judges. Since these cases were successful when there was no error the sole question for the Court is whether the conviction is generally unsafe based upon their own subjective assessment of all the facts of the case. As was shown by Model 1, the independent variables appear to be predictors of the answer to this question. Further analysis of the association between variables and the decision to quash convictions on the basis of fresh evidence is presented below.

⁸²⁸ Roberts (2009) 166.

7.8 Inclusion of the primary legal variable as an independent variable

In a final analysis, the outcome variable was reverted to the 'Was Appeal Allowed?' variable. This will be called Model 3. Model 1 was amended. All variables with a p-value higher than 0.05 were removed and the 'Did Error Occur?' variable added. As this is a study of the decision-making of the judges, it was decided to retain the three judge variables in this analysis. This meant that the dataset was its full size of 472 appeals, with 9 independent variables in the model.

As was expected, the inclusion of the legal variable significantly improved the performance of the model. The final model was statistically significant, with a p-value of less than 0.000. The Cox and Snell and Nagelkerke pseudo R² scores were 0.26 and 0.37 respectively. This is a considerable improvement upon the previous models presented in this chapter. The Hosmer and Lemeshow test was not significant (p = 0.266) meaning the model is a good fit for the data. The classification table is shown below. It can be observed, as would be expected, the overall correct classification percentage has increased, in particular for cases predicted to have been allowed.

Table 7.15: Classification Table: Model 3.

	Appeals Allowed	Appeals Dismissed	
Correctly Classified	75 (55.6%)	307 (91.1%)	
Overall Correct			80.9%)

The variables included in this model are shown below. Variables significant at the 5% level are marked with an asterisk.

Table 7.16: Variables in equation: Model 3.

	B	S.E.	Sig.	Exp(B)
Judge 1	1.092	0.546	0.046*	2.980
Judge 2	-0.825	0.615	0.180	0.438
Judge 3	0.764	0.390	0.050*	2.147
Bad Character	-0.879	0.260	0.001*	0.415
Denial of MR	0.377	0.260	0.148	1.458
Historical Offence	0.523	0.376	0.164	1.687
CCRC	1.230	0.411	0.003*	3.422
Single Judge	-0.537	0.270	0.047*	0.585
Did Error Occur	2.346	0.260	0.000*	10.441
Constant	-1.621	0.290	0.000	0.198

* = Significant at 5% level

The 95% confidence intervals for the variables' odds ratios is shown below.

Table 7.17: 95% confidence intervals for exp(B): Model 3.

Variable	Lower	Upper
Judge 1	1.022	8.692
Judge 2	0.131	1.463
Judge 3	1.001	4.609
Bad Character	0.249	0.692
Denial of MR	0.875	2.429
Historical Offence	0.808	3.523
CCRC	1.529	7.660
Single Judge	0.345	0.992
Did Error Occur	6.275	17.373

After this model was completed, all the original variables which were included in the purposive selection procedure were reintroduced into the model individually. The results of this closing analysis are not reported in this thesis, but none of the variables were statistically significant at the 5% level or had any discernible impact on the accuracy of the model, according to the R^2 statistics and classification matrix.

7.9 Summary of findings

These results are considered more closely in Chapter 8, but an initial impression of the results will now be provided. The variable representing the strongest measured relationship with the decision to allow an appeal, with an odds ratio of 10.441, is the 'Did Error Occur?' variable. This means that the odds of having a conviction quashed are measured to increase 10-fold if the Court finds an error occurred. Note that the broad confidence interval indicates that the true odds ratio may reach as high as 17.373. Whilst the confidence interval is broad, the lower bound reaches 6.275 meaning that the variable clearly has a strong relationship with the outcome of appeals. It has been argued that this variable is a counterweight to the factual and demographic variables, and is a measurement of the 'legal model' or the Court having decided appeals impartially in accordance with the law.

A number of the independent variables remain statistically significant predictors. Some of the judge variables remain statistically significant predictors of outcomes in combination with the legal variable. It is important, however, to notice the confidence intervals of the odds ratios. At the lower bound of the confidence interval for Judges 1 and 3, the odds ratio is only slightly above 1. The confidence interval is also broad, suggesting a lack of precision. Based upon this finding, there is not strong support for the claim that these judges had a clear relationship with the outcome of appeals.

The inclusion of a legal variable in the model adds significantly to the accuracy of the modelling. This suggests that studies which overlook the law may be overlooking a significant influence on Court of Appeal outcomes. Overall, the statistics of this model show that the 'Did Error Occur?' variable has by far the strongest association with the outcome of appeals amongst the variables which have been analysed. This does not mean that the Court has been proven to have determined appeals impartially, but the patterns found in this data add little support to the behavioural or attitudinal models of judicial decision-making.

The results of this study can best be summarised as indicating that there is insufficient evidence that the Court determined appeals other than in an impartial manner. As has been explained throughout this thesis, the methods employed in this study – a non-reactive observational study – mean there are limitations as to the conclusions which can be reached. How well the concept of impartiality has been modelled is considered in Chapter 8. The finding of statistical significance of some variables, and lack of significance of others, does not mean that the question of the Court's impartiality is resolved. Rather, as an initial exploration, it has been shown that some tentative patterns exist within the data and these are more indicative of the legal or institutional models.

7.10 Additional analysis of data – fresh evidence and CCRC referrals

Model 3, shown above, represents the final model of the whole dataset of murder and rape appeals analysed in this thesis. This is the primary model which will be discussed in Chapter 8. One difficulty with the above model is that it may miss the importance of fresh evidence and CCRC referrals on the outcome of appeals. As has been shown by the descriptive statistics relating to grounds of appeal, fresh evidence was the most frequently successful ground of appeal. Fresh evidence appeals may not be completely captured by the above model because they often do not contain procedural irregularities. Thus, while the 'Did Error Occur?' variable may measure the association between procedural irregularities and outcomes, it may not capture how the independent variables are associated with the admission of fresh evidence.

CCRC referrals are not 'ordinary' appeals against conviction, but are by their nature exceptional. Almost all CCRC referrals will have already been before the Court of Appeal and been rejected. CCRC referrals are made when the appellant continues to protest the conviction and lodges an application to the CCRC. The CCRC then conduct a review of the case and can refer the conviction back to the Court of Appeal. CCRC referrals more frequently uncover true miscarriages of justice, in the sense of the conviction of the innocent, in a way which is much rarer in routine appeals against conviction. Accordingly, they could be considered

as outliers in the sample – more likely to raise fresh evidence; have already had an extensive review; and more likely to uncover miscarriages of justice. Moreover, it is already known that a high proportion of CCRC appeals are successful.

In order to account for some of these factors, further analysis was conducted to supplement Model 3. As it is known that fresh evidence is a frequently successful ground of appeal, it is highly likely that some of the variation in outcomes is explained by the Court's reception of fresh evidence. In the data collection exercise, a variable was collected which was answered 'yes' if the appeal was quashed due to fresh evidence. Since this is itself an outcome variable, it could not be included as an independent variable because it would be collinear to the overall outcome variable. This would compromise the statistics because whenever the 'quashed due to fresh evidence' variable was answered 'yes', the overall outcome variable would also be 'yes', and this could confound or suppress other variables. One possibility was to conduct further binary logistic regression analyses with the 'quashed by fresh evidence' variable as the outcome, in a similar way as was done to the 'Did Error Occur?' variable. This was also problematic because the quashed by fresh evidence variable occurred on only 59 occasions, and so was a comparatively rare event. This makes it difficult to accurately model.

The preferred solution was to conduct analysis of the relationship between the independent variables and the 'quashed by fresh evidence' variable using Pearson chi-square (X^2) analysis. Chi-square analysis tests the null hypothesis that two variables are independent of each other. It does this by calculating the difference between the observed frequencies of convictions quashed by fresh evidence when the independent variable is present, with what would be expected if there was no relationship, or any relationship was due to chance.⁸²⁹ The greater the disparity between the observed and the expected frequencies, the less likely

⁸²⁹ G Norris, F Qureshi, D Howitt, D Cramer, *Introduction to Statistics with SPSS for Social Science* (Pearson 2012) 179.

it is that this null hypothesis is true.⁸³⁰ This allows for the analysis of the strength of a relationship between two categorical variables. This will allow for an analysis of whether the variables included in Model 3 have a relationship with the decision to quash a conviction on the basis of fresh evidence. The results of this analysis will then be considered in light of Model 3 to provide a more adequate analysis of the Court's decision-making and its impartiality.

All the variables included in Model 3, excluding the 'Did Error Occur' variable, were analysed in a 2x2 chi-square analysis. The following Table shows the p-values for each of the variables included in this analysis.

Table 7.18: Results of χ^2 Analysis.

Independent Variable	p-value
Judge 1 ⁸³¹	0.028*
Judge 2 ⁸³²	0.406
Judge 3	0.025*
Bad Character	0.010*
Denial of MR	0.913
Historical Offence	0.957
CCRC	0.000*
Single Judge	0.000*

* = significant at 5% level

Whilst Table 7.18 shows whether the variables are associated with each other, it gives little indication of the strength of the relationship. The strength of the correlation between each variable and the 'quashed by fresh evidence' variable was calculated using the phi-coefficient, which states the level and direction of any correlation. The higher the number, between +1 and -1, the stronger the correlation. A coefficient of 0.00 means there is no relationship. The following Table shows the phi-coefficients for the above variables.

⁸³⁰ *ibid.*

⁸³¹ This variable violated the assumption that there will be five expected values in each field. Accordingly Fisher's exact test p-value is provided.

⁸³² As per previous note.

Table 7.19: Phi-coefficients.

Independent Variable	Phi coefficient
Judge 1	0.111
Judge 2	-0.051
Judge 3	0.103
Bad Character	-0.119
Denial of MR	0.005
Historical Offence	-0.002
CCRC	0.342
Single Judge	-0.283

As would be expected, variables with larger p-values have extremely small phi coefficients, suggesting little relationship between the variable and the quashing of a conviction due to fresh evidence. It will be observed that most of the variables are only weakly positively or negatively associated with the decision to quash a conviction on the basis of fresh evidence. It is only CCRC references and the 'single judge' variables which have a correlation coefficient of higher than 0.200. The same variables as were statistically significant predictors in Model 3, are also statistically significant in the X^2 test. For these variables, the X^2 null hypothesis that the variables are independent of each other (there is no association) cannot be accepted. This suggests that as well as the variables being predictor of the outcome of appeals generally, there is some further relationship between the variables and the decision to quash convictions on the basis of fresh evidence. On the basis of the X^2 statistic, it is not possible to say what effect the variables have on the likelihood of having a conviction quashed on the basis of fresh evidence, as can be done in binary logistic regression analysis. The low phi-coefficients for these variables suggests that the relationship would be marginal, but the relationship is there.

As a final analysis of the dataset, the CCRC appeals were removed from the sample and a new binary logistic regression analysis was conducted. By removing the CCRC appeals, the issue that these appeals are 'outliers' is dissipated. With the CCRC referrals removed, the dataset was 428 cases, with 112 successful appeals (26.2%), and 316 unsuccessful appeals (73.8%).

Accordingly, this dataset has no EPV issues if the variables in Model 3 are utilised. The same variables as were included in Model 3 were used in a further binary logistic regression model, with the exception of the CCRC variable. The results are shown below.

The Cox and Snell and Nagelkerke pseudo R² statistics were 0.27 and 0.40, representing some improvement upon Model 3. The Hosmer and Lemeshow test was not significant (p = 0.585), meaning the model was a good fit for the data. The classification Table for this model is shown below.

Table 7.20: Classification table: Model 4.

	Appeals Allowed	Appeals Dismissed	
Correctly Classified	64 (57.1%)	292 (92.4%)	
Overall Correct			83.2%

This model also improved in classification accuracy when compared to Model 3, with overall correct classification increasing from 80.9% to 83.2%. Moreover, the percentage of appeals predicted to have been successful also showed an increase. The variables included in this model and their p-values is shown below.

Table 7.21: Variables in equation: Model 4.

	B	S.E.	Sig.	Exp(B)
Judge 1	0.938	0.614	0.126	2.556
Judge 2	-0.858	0.635	0.177	0.424
Judge 3	0.995	0.432	0.021*	2.705
Bad Character	-0.924	0.286	0.001*	0.397
Denial of MR	0.508	0.285	0.074	1.662
Historical Offence	0.444	0.406	0.274	1.559
Single Judge	-0.567	0.279	0.042*	0.567
Did Error Occur	2.597	0.280	0.000*	13.426
Constant	-1.772	0.310	0.000	0.170

* = significant at 5% level

When CCRC referrals are removed from the dataset, Judge 3, bad character, single judge, and Did Error Occur continue to be statistically significant predictors of outcomes. By this model, Judge 1 is now no longer a statistically significant predictor of outcomes. When CCRC appeals are removed, the odds ratio of the 'Did Error Occur?' variable increases to 13.426. The primary contribution of this model to the overall analysis of the Court may be that, in fact, the removal of the CCRC appeals from the sample did not greatly alter the results of Model 3. Whilst the classification accuracy was improved, it was based upon a smaller dataset and so was not a 'complete' picture of the Court's decision-making, in the way that Model 3 is. CCRC referrals may not, therefore, be as 'outlying' as was thought.

7.11 Conclusion

This study has been designed to capture some pertinent components of the principle of impartiality. It has been sought to test whether the Court appeared impartial, by exploring whether there is a relationship between a range of variables and the outcome of appeals against conviction. As was shown by the purposive selection procedure, most of the independent variables were weak predictors of outcomes in the Court of Appeal. Since there has been no previous study of this nature on decision-making in the Court of Appeal this does not confirm or refute previous studies, but provides some findings which can be confirmed or refuted in later studies. Similar to other ELS studies, a number of variables are shown to be statistically significant predictors of outcomes.

As is shown in Model 3, the variable with the strongest association with the outcome of appeals is the 'Did Error Occur?' variable. It has been explained that previous studies have found that the presence / absence of an error is related to the likelihood of a successful appeal. It has been claimed in this thesis that this factor can be understood as being indicative of the 'legal model' or impartial decision-making. Some of the behavioural factual and demographic variables analysed were associated with the outcome of appeals.

The analysis conducted and reported in this chapter was designed to test the concept of impartiality. The rationale was that the variables were observable implications of impartiality or a lack of it. The next chapter explores these results further. A key question addressed by Chapter 8 is how successful this study has been in capturing the concept of impartiality, and so what claims can be made regarding the impartiality of the Court of Appeal.

Chapter 8

Discussion: The Impartiality of the England and Wales Court of Appeal (Criminal Division)

Introduction

This thesis has critically analysed the creation of the England and Wales Court of Appeal (Criminal Division), its powers, and the perception of potential deficiencies in its decision-making. This study had analysed the Court from an Empirical Legal Studies (ELS) perspective. It did not address directly the question of how well the Court performs in correcting injustice, but instead sought to explore whether any of a range of variables are associated with the outcome of appeals against conviction. These are the 'observable implications' of impartiality or a lack of it, and this is how the concept of impartiality has been sought to be captured. It has been found that variables which are observable implications of impartial decision-making have the strongest relationship with outcomes. These variables measure the controlling role of law, or the Court's relationship with other institutions. Factual and demographic variables have a limited association with outcomes. The study presented in this thesis is an initial exploration of the question of the Court's impartiality. Thus, the patterns found in the data are consistent with the observable implications of having determined appeals in an impartial manner, but more evidence is needed to conclude this with certainty.

This chapter explores what the results of this study mean about the Court's decision-making. It will be discussed how successful this study has been as an exploration of the Court's impartiality, given the limitations of the methods employed. It will be explained that the results of this analysis, to a considerable extent, are both supported by, and support, the previous research on the Court. This thesis has, however, examined the Court's decision-making in a new and original way and this provides further insights into the Court's decisions. In Chapters 2 and 3, theories of judicial decision-making by Lord Bingham, Hart, Dworkin, and the Legal Realists were discussed. Conclusions will be offered,

based upon the reading of cases and the analysis of the data, as to which of these theories appears to be best reflected in the Court of Appeal judgments.

8.1 Overview of study

This thesis provides an empirical analysis of whether appellants, and other users and observers of the Court of Appeal, can be confident that the Court appeared to have determined appeals against conviction in an impartial manner. Impartial decision-making is the null hypothesis or default assumption of this study, which could only be rejected if there is sufficient evidence. This is because impartiality is inherent within the 'legal model' of judicial decision-making. The 'legal model' postulates that judicial decisions are primarily determined by judges impartially applying the law, and so judicial discretion is constrained by law. In Britain, as there is no great tradition of ELS on judges, there was no empirical *a priori* reason to suspect that judges do not adhere to the legal model. It was found that, whilst some variables potentially indicative of behavioural or attitudinal models were associated with the outcome of appeals against conviction, it is unclear how strong this association is. Moreover, variables which were offered as being observable implications of the legal model and impartial decision-making were found to be strongly associated with outcomes. Accordingly, there is insufficient evidence to reject the legal model.

The meaning of impartiality, and its importance, was discussed in Chapter 2. To address the question of the Court's impartiality, a sample of 472 appeals against conviction was created, which recorded the demographic, factual, and legal independent variables collected from each case. The analysis began with a replication of Roberts's,⁸³³ and Heaton's⁸³⁴ studies of the grounds of appeal argued in appeals against conviction. In each of these earlier studies, data relating to the grounds of appeal was collected in order to assess the Court's

⁸³³ S Roberts 'The Decision-Making Process of Appeals Against Conviction in the Court of Appeal (Criminal Division) (DPhil Thesis, London School of Economics and Political Science, 2009). (Hereafter Roberts (2009)).

⁸³⁴ SJ Heaton, *A Critical Evaluation of Using Innocence as a Criterion in the Post-Conviction Process* (DPhil Thesis, University of East Anglia, 2013). (Hereafter Heaton (2013)).

approach to determining appeals. Roberts read the first 300 available appeals from 2002, and Heaton read all Court of Appeal cases decided in the calendar year 2009, all CCRC referral appeals, and consulted CCRC case files. In this thesis, all available murder and rape appeals against conviction decided between January 2006 and December 2010 were collected. The intention behind the initial analysis of grounds of appeal was to discover whether these two specific offences raised different grounds of appeal to those seen in the previous studies, which included not only murder and rape appeals but appeals relating to numerous different crimes. It was shown in Chapter 7 that there were significant similarities between these three studies, with the important exception of fresh evidence. The implications of this finding are discussed later in this chapter.

The method of data collection in the previous studies and the present were identical. The cases were downloaded from legal databases and analysed by means of a case-by-case quantitative content analysis before being coded into SPSS. The difference between this thesis and the earlier studies was that as well as the collection of grounds of appeal, a wider variety of other variables were also collected. Thus, although there are many similarities in methodology and subject matter between this thesis and the earlier studies, they diverge at the point at which this thesis moves beyond grounds of appeal and into considering factual, demographic, and other legal variables as predictors of outcomes. There has been no known similar study of this nature conducted on the Court of Appeal. It is this difference which leads to the original methodology and original analysis of decision-making contained in this thesis. In this thesis, the decision-making of the Court was analysed using a series of binary logistic regression analyses, to determine the extent of any statistical relationships between success in the Court of Appeal and the independent variables.

The analysis was in the form of a series of null hypothesis significance tests. It was determined whether each independent variable was a statistically significant predictor of a successful appeal by considering a number of statistics, including the p-value of each variable, and the overall predictive accuracy of each model

using pseudo R^2 statistics, classification tables, confidence intervals and goodness of fit tests. A p-value is a percentage between 0 and 1 which quantifies the degree to which the data confirms with the null hypothesis prediction. Smaller p-values can indicate that there is a greater statistical incompatibility between the data and what the null hypothesis would predict, which may therefore give some reason to doubt the null hypothesis. The standard values for rejection of the null hypothesis in social science research are $p = <0.05$ and $p = <0.01$, respectively relating to the 5% and 1% levels of significance.

The use of p-values was discussed in Chapter 5. It was shown that the use of p-values is controversial, as p-values are frequently misinterpreted. It is important to recall that the 'rejection' of a null hypothesis does not mean that the alternative hypothesis is true. Furthermore, the rule that significance is reached if the p-value is smaller than 0.05 is an arbitrary figure. As was discussed in Chapter 5, the limitations of significance testing mean that they provide only tenuous evidence against a null hypothesis. The results of this study are returned to below, where the strength of the findings regarding the impartiality of the Court in light of this, and other, limitations of the research design are discussed.

This thesis has eschewed examination of the normative significance of particular cases, or the correctness of particular cases, in order to instead consider whether any factors are associated with Court's decisions. The aim of the study was to conduct an exploration of whether statistical patterns exist between the outcome of appeals against conviction and the independent variables. This is an ELS perspective to legal scholarship, in that it sought to discover with greater objectivity what occurs in the Court of Appeal. This is in contrast with some previous studies of the Court of Appeal, which have examined the Court's decision-making from a 'miscarriage of justice' perspective. The thesis has considered the methodologies of the many empirical studies of judicial decision-making primarily from the United States.

The infamous case of Timothy Evans highlights the different approaches adopted in this thesis and previous studies.⁸³⁵ Timothy Evans was convicted of the murder of his wife and child. He appealed in 1950, but the appeal was dismissed, and he was duly hanged. It is now believed, beyond reasonable doubt, that he was innocent of the crime. It was later revealed that his landlord was a serial killer, and he confessed to the murders. The decision of the Court of Appeal to dismiss his appeal in 1950 was wrong, and arguably restrictive. The Court of Appeal did not know that Evans was innocent when they dismissed the appeal, and they did not know at the time that his landlord was a serial killer, but the decision to dismiss the appeal was wrong. Evans's appeal against conviction could be criticised on other grounds. Counsel for the Crown, Christmas Humphreys, was appearing before his own father, also Christmas Humphreys, in the appeal. This would never be acceptable in the modern day, as the appearance of a lack of impartiality would be too great. The decision could therefore be considered illegitimate, regardless of whether it was right or wrong. It is this aspect of the Court's decision-making, its legitimacy to reach decisions, with which this thesis has been concerned.

8.2 Grounds of appeal argued in the Court of Appeal in murder and rape appeals – Is the Court of Appeal restrictive?

Since the 1970s, there have been four empirical studies of the grounds of appeal raised by appellants. These four studies are by Knight,⁸³⁶ Malleon,⁸³⁷ Roberts,⁸³⁸ and Heaton.⁸³⁹ The latter two are more important to this thesis because only they were conducted after the enactment of the *Criminal Appeal Act 1995*, and so the older studies have been superseded by the more recent research. Their findings in relation to grounds of appeal were critically analysed in Chapter 7.

⁸³⁵ See *R v Evans* (1950) 34 Cr App R 72.

⁸³⁶ M Knight, *Criminal Appeals* (Stevens and Sons 1970).

⁸³⁷ K Malleon, *The Royal Commission on Criminal Justice Review of the Appeal Process*, Research Study No 17 (HMSO 1993).

⁸³⁸ S Roberts, 'The Royal Commission on Criminal Justice and Factual Innocence: Remedying Wrongful Convictions in the Court of Appeal' (2004) (1)(2) JJ 86; and Roberts (2009).

⁸³⁹ See Heaton (2013).

In Roberts's, Heaton's, and this study, two grounds of appeal dominated the arguments raised by appellants: complaints regarding the trial judge's summing up, and allegations that the trial judge incorrectly exercised his evidential discretion. The results of this study therefore partially support the findings of previous studies, as procedural irregularities do form the majority of the Court's workload in murder and rape appeals. At least one of these two grounds of appeal was raised in 80% of appeals against conviction in murder and rape appeals, and they together accounted for almost 45% of all grounds of appeal. Therefore, murder and rape appeals are just as likely to raise procedural irregularity appeals as was found in earlier studies which considered a range of different offences. Roberts and Heaton also found that fresh evidence and lurking doubt appeals were rare and unlikely to be successful. In relation to lurking doubt appeals, this thesis confirms this previous finding. In this sense, murder and rape appeals appear to be similar to most other appeals in the Court.

Fresh evidence, however, was raised more frequently in this study than was seen in either Roberts's or in Heaton's studies. In contrast to their findings, fresh evidence was raised in 130 appeals, which was 27% of all cases and 15% of total grounds. Fresh evidence was successful on 59 occasions, which was 45% of the cases where it was raised, and was the most frequently successful individual ground of appeal. The corresponding fresh evidence success rate was 25% for Roberts; and 31% for Heaton. This suggests that fresh evidence is more likely to be raised, and is more likely to be a successful ground of appeal in murder and rape cases than it is in the Court of Appeal generally. Moreover, as was shown in Chapter 7, whilst CCRC referrals did account for a large number of the CCRC referrals, murder and rape non-CCRC appeals still raised fresh evidence more commonly than seen in earlier studies, and it was more likely to be successful.

The low success rate of fresh evidence seen in the earlier analyses is seen as one piece of evidence that the Court is restrictive or fails to correct miscarriages

of justice.⁸⁴⁰ This is because fresh evidence most clearly raises the possibility of innocence, and requires the Court to go beyond its comfort zone, and come close to retrying the appellant on the basis of the fresh evidence. The conclusion drawn from this element of the thesis, that fresh evidence was actually the most frequently successful ground of appeal, and was successful 45% of the times it was raised, must therefore lead to some pause for thought regarding the conclusion of restrictiveness. It could also be considered one of the most significant findings of this study.

It appears that murder and rape convictions are more susceptible to finding fresh evidence, and to being rendered unsafe by fresh evidence. There are some plausible reasons why this may be the case. Murder trials often involve scientific evidence, frequently at the cutting edge of pathology, and such evidence is always liable to be undermined or challenged by later developments. The experts giving evidence may offer exaggerated, or simply wrong, opinions which may be picked up by other experts or new lawyers. Professor Sir Roy Meadow's evidence, implicated in a number of 'shaken baby' cases,⁸⁴¹ is one example of this occurring. His evidence was described by the Court of Appeal as being grossly negligent.⁸⁴² A further example is former Home Office pathologist Dr Michael Heath, who was referred to the Home Office pathology regulators by his own colleagues.⁸⁴³ His methodological errors in attributing causes of death was a source of a number of fresh evidence appeals against conviction.⁸⁴⁴ There were numerous appeals against convictions for murder where fresh evidence provided evidence of diminished responsibility due to previously undiagnosed (or incorrectly diagnosed at trial) mental illness.⁸⁴⁵ Scientific evidence given at trial can be rendered invalid or suspect if scientific knowledge changes, and this new

⁸⁴⁰ S Roberts, 'The Royal Commission on Criminal Justice and Factual Innocence: Remedying Wrongful Convictions in the Court of Appeal' (2004) (1)(2) JJ 86.

⁸⁴¹ Eg *R v Sally Clark* [2003] EWCA Crim 1020; *R v Angela Cannings* [2004] EWCA Crim 1.

⁸⁴² *GMC v Professor Sir Roy Meadow v HM Attorney General* [2006] EWCA Civ 1390 [82]

⁸⁴³ *R (oao Heath) v Home Office Policy and Advisory Board for Forensic Pathology* [2005] EWCA Crim 1793.

⁸⁴⁴ See *R v Puaca* [2005] EWCA Crim 3001; *R v O'Leary* [2006] EWCA Crim 3222, *R v Khokhar* [2007] EWCA Crim 1756; *R v Stanley* [2008] EWCA Crim 603.

⁸⁴⁵ Eg *R v Erskine*; *R v Williams* [2009] EWCA Crim 1425.

knowledge can be adduced at appeal as fresh evidence.⁸⁴⁶ All of these factors may mean that fresh evidence may be more likely to arise in murder appeals than in other offences.

In rape trials, especially historical rape trials, the only evidence called at trial is often medical evidence and / or the complainant's testimony. Both of these are susceptible to be dislodged by fresh evidence. Changing understanding of, or rather, appeals challenging the accepted understanding of, injuries caused by childhood sexual abuse was a potent source of fresh evidence appeals in rape.⁸⁴⁷ Fresh evidence was frequently raised which sought to challenge the integrity of the complainant's testimony, by, for example, demonstrating that the complainant made a claim for compensation which allegedly exaggerated the offences, or other attacks on credibility.⁸⁴⁸ Fresh evidence did arise where the complainant(s) admitted that the allegations were false.⁸⁴⁹ These kind of attacks on credibility on the basis of fresh evidence are difficult to envisage for many other offences which the Court hears.

In the absence of further research or data, it is not known whether only certain kinds of offences, in addition to murder and rape, form the bulk of fresh evidence appeals. Further research could be conducted to discover this. If the majority of appeals in Roberts's and Heaton's samples are by their nature not generally susceptible to uncovering fresh evidence or to being quashed by fresh evidence, this might explain why they saw that fresh evidence was raised relatively infrequently. This might undermine the claim that the Court is restrictive or overly cautious about fresh evidence. Certainly, the finding of this thesis that fresh evidence was successful 45% of the times it was raised, or 39% of the times it

⁸⁴⁶ Eg *R v Reed and Reed*; *R v Garmson* [2009] EWCA Crim 2698 in which the Court considered the state of knowledge relating to DNA evidence. See also *R v Lawless* [2009] EWCA Crim 1308, relating to new understanding of how abnormally suggestible suspects may wrongly confess to crimes.

⁸⁴⁷ Eg *R v Michael A* [2006] EWCA Crim 905; *R v Martin T* [2008] EWCA Crim 3229; *R v PF* [2009] EWCA Crim 1086.

⁸⁴⁸ Eg *R v B* [2008] EWCA Crim 559; *R v Charova* [2008] EWCA Crim 1767.

⁸⁴⁹ Eg. *R v Barker* [2006] EWCA Crim 3249; *R v Anwar* [2007] EWCA Crim 3226.

was raised in non-CCRC referral cases, does not support a conclusion of restrictiveness in relation to fresh evidence.

The Court of Appeal explained in *R v Erskine*,⁸⁵⁰ that the primary question for the Court when deciding whether to receive fresh evidence is whether it is in the interests of justice.⁸⁵¹ The Court may be called restrictive for declining to receive fresh evidence by interpreting the section 23 requirements in a narrow way. The first two tests under section 23, whether the evidence appears 'capable of belief' and whether the Court thinks the evidence 'affords any ground for allowing the appeal', are tests of the Court's view of the fresh evidence. To be more liberal in applying these tests, the Court would need to receive fresh evidence if it finds the evidence not capable of belief, or where it does not think the evidence makes the conviction unsafe. It would be an unusual situation for the Court to quash convictions on the basis of fresh evidence which it does not itself believe, as the Court made clear in cases such as *Stafford v DPP*⁸⁵² and *R v Pendleton*.⁸⁵³

Heaton argued that the Court 'displays confusion about the correct test to be applied in fresh evidence cases and applies the test in an unpredictable manner'.⁸⁵⁴ On the contrary, this thesis concludes that the Court is sufficiently clear in the test to be applied, and that the unpredictable manner of decision-making is due to the different facts of cases. As was explained in Chapter 4, in fresh evidence appeals, the question for the Court is whether they think the conviction is unsafe, and one way of judging that is to consider whether it is sure that the jury would have still convicted if it had heard the fresh evidence.⁸⁵⁵ As Ashworth and Redmayne say, there is relatively little at stake as to whether the 'correct test' is the 'jury impact' test or the subjective 'judge impact' test, because the question is whether the conviction is safe or unsafe.⁸⁵⁶ Different cases, or

⁸⁵⁰ [2009] EWCA Crim 1425.

⁸⁵¹ *ibid* [39].

⁸⁵² [1974] AC 878

⁸⁵³ [2001] UKHL 56

⁸⁵⁴ Heaton (2013) 175.

⁸⁵⁵ See *R v Pendleton* (n 853).

⁸⁵⁶ A Ashworth and M Redmayne, *The Criminal Process* (4th edition, Oxford University Press 2010).

different judges, may appear to be applying one test rather than the other, but they in effect amount to the same thing – does the Court think the conviction can stand in light of the fresh evidence? There are numerous examples of judges stating this, but one is sufficient to demonstrate the point. In *R v Williams*⁸⁵⁷ Auld LJ said:

‘Having received ... “new” psychiatric evidence ... a court may allow an appeal against a conviction for murder as unsafe on the ground that, if that evidence had been before the trial jury, they might have acquitted the appellant on one or other of the defences of diminished responsibility, or provocation where applicable, or self-defence’.⁸⁵⁸

In that case, the Court of Appeal concluded that the conviction was safe because the material upon which fresh expert evidence was based was of so little weight, and, in fact, potentially damaging to the appellant’s case, it could not have changed the jury verdict. Heaton’s claim that the Court uses the section 23 provisions in order to restrictively dismiss appeals which had already been decided on some ‘underlying, unarticulated basis’⁸⁵⁹ suggests that there is some insidious aspect to the Court’s decision-making. An alternative explanation is that the ‘unarticulated basis’ of appeals is that the Court simply does not believe the evidence, or does not think that the evidence would have made a difference to the verdict; i.e. the Court is sure of guilt.

If the Court dismisses an appeal, this does not necessarily mean the Court is restrictive or slow to act, but instead may demonstrate that there was no basis in fact or law to find the conviction unsafe at the time the case is before the Court. As a result of this, the Court might on occasion uphold the conviction of a person who is actually innocent. This would of course be a miscarriage of justice, and devastating to the victim of the miscarriage of justice. However, unless the Court is to be expected to quash convictions whenever there is any doubt about guilt,

⁸⁵⁷ [2007] EWCA Crim 2264.

⁸⁵⁸ *ibid.*

⁸⁵⁹ Heaton (2013) 211.

which will almost always arise, there appears to be no way to avoid the situation that the Court may on occasion fail to correct the conviction of the innocent. The best way to avoid the conviction of the innocent, it is argued, may be to make further efforts to prevent the conviction of the innocent occurring, beginning by reforming police investigation practices.

The allegation of restrictiveness, it is submitted, does not appear warranted based upon the empirical analysis of this study and a more neutral analysis of the Court's decision-making. As explained in Chapter 4, previous analysis of the Court has been conducted from a miscarriage of justice perspective. Naughton, for instance, criticises the Court of Appeal as he felt that it does not properly perform its function of 'assisting potentially factually innocent victims to have miscarriages of justice overturned.'⁸⁶⁰ From the ELS perspective of this study, the conclusion that the Court's approach is 'restrictive' does not appear stable or objective. There will always be appeals which are unsuccessful, and there may always be doubt whether those decisions were correct or whether the appellant is innocent. However, having doubt about a particular decision cannot be evidence of a restrictive approach, because that is not empirical evidence.

The allegation of a restrictive approach implies that quashing more convictions by being more liberal would produce more justice. But, as Roberts acknowledged, the Court cannot quash every conviction, and it must find a way to determine which appeals can stand and which cannot.⁸⁶¹ Whilst Naughton is highly critical of the Court's decision-making processes,⁸⁶² he provides no evidence that the Court's approach systematically has led to it upholding convictions which should have been quashed. Moreover, it is unclear how such evidence could be provided, as it is rarely ever known which appellants, if any, were actually innocent.

⁸⁶⁰ M Naughton, *The Innocent and the Criminal Justice System: A Sociological Analysis of Miscarriages of Justice* (Palgrave Macmillan 2013) 147.

⁸⁶¹ See Roberts (2009) 41-2.

⁸⁶² See Naughton (n 860) especially Chapter 6.

In this study, there were 337 appeals which were unsuccessful. It is possible that at least some of these appellants were in fact innocent. At least one appellant in the sample later had his conviction for murder quashed following a CCRC referral.⁸⁶³ Does this mean that all the unsuccessful appeals were potentially restrictive interpretations? Proponents of the allegation of a restrictive approach would need to explain which elements of the decision-making in these 337 cases are indicators of a restrictive approach, and which would be considered acceptable decision-making by the Court. It is difficult to locate such an indicator, because the Court's decision-making is focussed upon operating the unsafety test and it will quash all, but only, those convictions which it thinks are unsafe. Others may disagree with the Court's decision in the unsuccessful appeals, but, as Roberts noted, this does not mean that the conviction should have been quashed.⁸⁶⁴

The greatest difficulty with the allegation of a restrictive approach is that it is based upon the assumption that the approach leads to some appeals to be wrongly decided. There is no acceptable evidence that this routinely does occur. This thesis has eschewed the question of whether certain cases were rightly or wrongly decided, for the reason that it is very difficult to know whether the outcome in any particular appeal was wrong. Instead, whilst certain decisions have been discussed for their qualitative significance, this thesis has taken a broader approach to analysing the Court. It has sought to analyse the emergence of patterns in the data, taking a wider view of the Court's decisions. The question addressed was whether the Court appeared to have determined appeals in an impartial manner. The remainder of this chapter addresses that question.

8.3 Analysing the impartiality of the Court of Appeal

In addition to the analysis of the grounds of appeal raised in the murder and rape appeals against conviction, this thesis sought to analyse an additional component

⁸⁶³ Sam Hallam had his appeal dismissed in [2007] EWCA Crim 966, and quashed in [2012] EWCA Crim 1158.

⁸⁶⁴ Roberts (2009) 213.

of the Court of Appeal's decision-making. A variety of factual, demographic, and legal details were collected from each case, some of which formed the independent variables which were analysed. The methodology employed in this study has been explained earlier in this thesis, in depth. This methodology was important, because it explained how far the results were valid, leading to meaningful results. The results of this analysis forms an original contribution to the understanding of decision-making in the Court of Appeal. Furthermore, the methodology employed in conducting the research represents an additional original contribution.

The results of this element of the study have been presented in Chapter 7. In the final binary logistic regression model, Model 3, which included the 'Did Error Occur?' variable, six of the independent variables were statistically significant predictors at the 5% level of success or failure of an appeal against conviction in the Court of Appeal. These were: Judge 1, Judge 3, Bad Character, CCRC, Single Judge, and Did Error Occur? (the 'legal variable'). As has been explained previously, the rejection of these null hypotheses based on statistical significance is descriptive, and not inferential. This means that the rejection of the null hypotheses does not mean that the Court lacked impartiality. This is owing to a number of substantial limitations, which are explored in the next section.

The legal variable, 'Did Error Occur?' was the strongest predictor of successful appeals, with an odds ratio 10.441. This increased to 13.426 when CCRC appeals were removed from the analysis. This means that the odds of having a conviction quashed increase ten-to-thirteen fold if the Court finds an error occurred at trial. Moreover, the confidence interval for this variable was broad, with a lower bound of 6.275, suggesting that as a minimum appeals were 6 times more likely to be successful when an error occurred. The inclusion of the legal variable considerably improved the predictive power and accuracy of the models. In many studies of judicial decision-making, in particular 'attitudinal' studies of US Courts, there is rarely any measurement of law. This is a problem because the

attitudinal model argues against the 'legal model' claim that the law is a constraint on judicial discretion.⁸⁶⁵

The results of the binary logistic regression analyses conducted in this thesis revealed that some of the judges who heard more than 20 appeals were statistically significant predictors of the outcomes of appeals. This effect persisted when the legal variable was included in the analysis. If an appellant had Judge 1 or Judge 3 as a judge the appellant was, respectively, 2.147 and 2.980 times more likely to be successful. The p-values were significant at 0.05 and 0.04 respectively. As has been stated previously, this finding is tenuous and so provides no evidence that these judges, or the Court generally, lacked impartiality. This is compounded by the observation that the lower bound of the confidence intervals for these variables were very close to 1, suggesting only a weak association. As an initial exploration of the Court's impartiality, this study has uncovered a relationship between certain judges and the outcome of appeals against conviction. There is insufficient evidence to draw any further inference of a lack of impartiality from this data, but the observation of this emerging pattern would warrant further study. Further research could be conducted to determine whether this could be evidence of attitudinal decision-making, or whether this finding is a relic of institutional factors owing to the manner of case allocation in the Court.

The importance of the judge variables for this thesis is that the performance of the Court under the 'unsafety test', and, indeed, under all the previous tests, has been heavily dependent upon how particular judges have exercised their powers. It is possible to induce a number of possibilities for the finding of a statistical relationship between judges and the outcome of appeals. Judges 1 and 3 may be more likely to quash convictions when an error occurred because their own conceptions of justice leads them to think that convictions ought to be quashed when errors occurred at trials. Rather than this being evidence of a lack of

⁸⁶⁵ See JA Segal and HJ Spaeth, *The Supreme Court and the Attitudinal Model Revisited*, (Cambridge University Press 2002).

impartiality, these may judges demonstrate greater fidelity to law; the opposite of a lack of impartiality. The two-stage decision-making process in procedural irregularity appeals is important here. It was seen in Chapter 7 that often when the Court finds that an error occurred they will allow the appeal, but the Court also frequently dismisses the appeal. The finding that Judges 1 and 3 were predictors of successful appeals might suggest that they are more likely to take the next step and decide that errors make convictions unsafe.

The decision-making of the Court of Appeal in relation to the murder and rape appeals against conviction in the sample could be summarised in the following way. When the judges concluded that an error did not occur at trial, the appeal was usually dismissed. There were 278 appeals, out of a total of 472 appeals, which were dismissed when the Court found that no error had occurred. This adds support to, and is supported by, previous research into the Court of Appeal. It suggests that if an appellant cannot point to a procedural irregularity, the appeal is likely to be unsuccessful. It confirms Roberts's conclusion that the most common result if the Court finds that an error did not occur is that it will dismiss the appeal.⁸⁶⁶ One possible interpretation of this finding is that the Court of Appeal is restrictive, in that it will not normally quash convictions if it does not feel that an error occurred at trial. As explained above, this conclusion does not appear supported by this finding. An alternative interpretation, based on theories of judicial decision-making, is offered below.

Whilst in the majority of appeals in which no error occurred were dismissed, there were 49 appeals which were successful when the Court found that no procedural irregularity occurred during the trial. All but one of these was successful due to fresh evidence. This potentially adds further support to the concern that if an appellant reaches the Court with no fresh evidence or procedural irregularity there is hardly any chance of having the conviction quashed. The appeal which was successful when the Court allowed the appeal when there was no error at trial,

⁸⁶⁶ Roberts (2009) 149.

and no fresh evidence, was *R v Haigh*.⁸⁶⁷ She submitted three grounds of appeal: that the judge wrongly admitted certain evidence; that the judge should have acceded to a claim of no case to answer; and, simply, ‘the conviction for murder, rather than manslaughter, is unsafe’.⁸⁶⁸ The Court rejected the claim that the evidence should not have been admitted, and held that there was a case to answer.⁸⁶⁹ They allowed the appeal, and substituted a verdict of manslaughter, because:

‘There was no evidence at all on the basis of which the jury could reasonably decide whether the appellant had the intent to kill or cause really serious harm ... or the lesser intent which was sufficient for manslaughter ... we are driven to the conclusion that the murder conviction is unsafe’.⁸⁷⁰

It is pertinent to note that at no point did the Court directly or indirectly refer to the concept of ‘lurking doubt’, yet the case appears to be an archetype ‘lurking doubt’ case, in which the Court allowed the appeal because of concern that an injustice could have been done.

As was described in Chapter 7, as further analysis, it was sought to explore the relationship between the independent variables and the decision to quash convictions on the basis of fresh evidence. Using a chi-square analysis, it was shown that the same independent variables as were statistically significant predictors in Model 3, were weakly correlated individually with the quashing of convictions on the basis of fresh evidence. This suggests that the variables have a consistent relationship with the decision to allow appeals, either on the grounds of procedural irregularities or fresh evidence, but that the substantive importance of the variables is relatively low. When the chi-square analysis is considered alongside Model 3, it can be seen that the variables are individually associated with the decision to quash convictions on the basis of fresh evidence, and

⁸⁶⁷ [2010] EWCA Crim 90.

⁸⁶⁸ *ibid* [1].

⁸⁶⁹ *ibid* [23] and [92].

⁸⁷⁰ *ibid* [97].

collectively associated with the general question of the unsafety of convictions. However, this additional analysis suggested that the correlation between the variables and the decision to quash convictions on the basis of fresh evidence was low.

One significant finding from this study is that 59 appeals were unsuccessful, and 86 appeals were successful, when the Court concluded that an error had occurred at trial. The binary logistic regression analysis showed that there is a tenfold increase in the odds of a successful appeal when the Court found that an error occurred, in combination with the other variables included in the model. This is a large odds ratio, demonstrating a strong relationship between this variable and the outcome of appeals against conviction. The observation that 59 appeals were dismissed when the Court found that an error occurred suggests that this variable is not fully determinate of appeals. It suggests the presence of a gap or discretion when the Court finds that an error occurred. The 'two-step' process of decision-making in relation to the 'unsafety test' was discussed in Chapter 3, and it would appear that the observation of 59 unsuccessful appeals is an artefact of this process.

The CCRC variable was also a statistically significant predictor of success. This variable had an odds ratio of 3.422 meaning that appeals brought by the CCRC were more than three times more likely to be successful when considered in combination with the other variables included in the model. The finding that CCRC referrals are significant predictors of success is as expected because the CCRC can only refer appeals back to the Court of Appeal if they think it has a 'real possibility' of success. Moreover, appeals go through a detailed review by the CCRC before being referred. The CCRC's own statistics show that they have a success rate of cases referred of 65%.⁸⁷¹ Owing to this high success rate, the CCRC variable was always likely to be a positive predictor of outcomes. The CCRC variable was offered as an additional counterweight to the factual and

⁸⁷¹ <http://www.ccrc.gov.uk/case-statistics/> <6 June 2016>.

demographic variables included in this study. It was argued that this variable reflected the broader institutional relationships between the CCRC and Court of Appeal. In this instance, it would appear that the Court of Appeal and the CCRC are likely to draw upon each other. The CCRC must consider how it thinks the Court will respond to the appeal, and the Court will consider carefully that an independent body thinks that the conviction is potentially unsafe.

This finding relating to the CCRC variable is in contrast with the finding relating to the Single Judge variable. Owing to its negative coefficient, if leave was granted by the single judge there were reduced odds of having a conviction quashed. The odds of success if leave was granted by the single judge was 0.585 that of an appeal granted leave by other means, and this was statistically significant. This may suggest that appeals granted leave by the single judge are particularly susceptible to being dismissed. It should be noted that a 'Full Court' variable was collected, for appeals which were granted leave by the Full Court, but this was shown to be a weak predictor in the purposive selection procedure and so not carried forward to the next stage of analysis. There was a persistent statistical relationship between the single judge variable and unsuccessful appeals throughout the analysis. This may suggest that the Court of Appeal judges do not see the granting of leave by the single judge as a *prima facie* indicator of unsafety. This is in contrast to the CCRC variable, in which there must be a strong likelihood that the conviction will be considered unsafe, owing to the extensive review by the CCRC.

Owing to its negative coefficient, the question of whether the appellant had 'bad character', was a predictor of reduced odds of success. This variable required some interpretation in the collection process. The variable was always answered 'Yes' if the Court specifically referred to relevant bad character or previous convictions, and it was answered 'No' if the Court was silent on character, or if the transcript stated good character. However, at times, bad character was imputed, for instance if the appellant was a known gang member, or involved in drugs etc., notwithstanding that he may not have any convictions. This approach

is justifiable as the definition of bad character includes 'reprehensible conduct', which is interpreted broadly.⁸⁷² This variable may not, therefore, be a precise measure whether the appellants had 'bad character', and all observers might not agree with the coding decisions made. There were 210 cases coded as having bad character. If the appellant had bad character, as defined, the odds of a successful appeal were 0.415 that of an appellant without bad character, i.e. less than half as likely to be successful. As an attitudinal matter, it makes intuitive sense that appellants with bad character may be considered more likely to be guilty by the Court. Indeed, in limited circumstances bad character evidence can be admitted as evidence towards a defendant being more likely to have committed the crime with which he is charged. There is strong evidence that people convicted of a crime are more likely to commit more crimes.⁸⁷³

Despite this, the admission of bad character evidence in criminal trials can be criticised as propagating stereotypes or prejudice.⁸⁷⁴ Impartiality, or the 'legal model', maintains that judges ought to determine cases only with reference to the law, and not by bias or prejudice. With a p-value of 0.003, the bad character variable had the strongest association with outcomes after the 'Did Error Occur?' variable. This finding may suggest that previous offending behaviour creates a heavy burden for an appellant to overcome. This finding is not strong evidence of a lack of impartiality, however. As has been explained previously, and will be summarised in the next section, the rejection of null hypotheses on the basis of p-values does not mean that the alternative hypothesis is true. However, the emergence of this pattern in the data is one that may warrant further analysis in the future.

This section has analysed in more depth the results of Model 3. The intention behind this thesis was to collect a range of variables in order to explore the

⁸⁷² See Criminal Justice Act 2003, s 112, and Code for Crown Prosecutors guidance on bad character evidence http://www.cps.gov.uk/legal/a_to_c/bad_character_evidence/ (accessed 24/06/2017)

⁸⁷³ See M Redmayne, *Character in the Criminal Trial* (Oxford University Press 2015) 17-25.

⁸⁷⁴ *ibid*, Chapter 3.

Court's decision-making. It was intended that the variables would capture in part the principle of impartiality, to allow for conclusions to be drawn regarding whether the Court appeared to have determined appeals in an impartial manner. Accordingly, how well this study has captured the concept of impartiality is at the core of what kinds of claims can be made about the Court's impartiality. The next section discusses how successful this thesis has been at exploring the impartiality of the Court.

8.4 How successful is the study as an exploration of impartiality?

As an exploration of whether a range of variables act as predictors of decision-making in the Court of Appeal, this study has been successful. This study has shown that there are some observable patterns between the variables selected for analysis and the outcome of appeals. Owing to the limitations of the methods of this study, however, it is less clear whether it can be said that the impartiality of the Court has been modelled very successfully. It is clear that the measurement of impartiality is incomplete. This means that whilst it has been shown that there appears to be patterns in the data, the relationship between these patterns and the concept of impartiality is somewhat tenuous. Making an extraordinary claim about a Court – such as it lacks impartiality – requires extraordinary evidence, and the onus is always upon the researcher to provide such evidence. In this study, the evidence of a lack of impartiality is too limited to support such claims. The reasons for this are now recounted.

The analysis of the impartiality of the Court focussed upon the collection of a range of variables which were then analysed statistically for the emergence of patterns in the data. This kind of research is correlative, and as such causal inferences cannot be drawn. As was discussed in Chapter 5, the ideal conditions to be able to begin to make causal inferences is a randomised controlled experiment using validated (by replication) measures. This study, a correlative, non-reactive observational study, is several stages removed from this ideal condition. In an experiment, the researcher can control variables in order to observe whether changes in one variable lead to different outcomes when all the

other variables remain the same. This study has relied upon the observation that certain variables have an association with outcomes. This necessarily limits significantly how far it can reasonably be inferred that the variables have any actual effect on outcomes.

Epstein and King made this same point regarding ELS in 2002.⁸⁷⁵ They note that whilst frequently the aim of ELS is to draw inferences regarding the causal effect of one variable upon another, it may not be possible to draw causal inferences owing to how far removed ELS is from a genuine experiment.⁸⁷⁶ This is particularly true for this study, given its explorative nature and the use of some variables which have not been tested previously in earlier studies. This is one reason why the results of this study must be read cautiously and not misinterpreted. The observation that there appears to be a statistical relationship between some variables and outcomes, cannot be conclusive of whether there exists a true causal relationship. It is important for the future credibility of ELS that researchers carefully stipulate the epistemological limits of the method.

A related point is the method of testing for statistical relationships. As was discussed in Chapter 5, the American Statistical Association (ASA) have recently produced a report on p-values and statistical significance.⁸⁷⁷ The ASA statement followed growing concern of a 'replication crisis' in numerous branches of science.⁸⁷⁸ A key concern of the ASA was the interpretation of p-values and 'statistical significance'. It was explained in Chapter 5 that rejecting a null hypothesis due to statistical significance is not evidence that the alternative hypothesis is true. Moreover, the concept of 'statistical significance' if a p-value is smaller than 0.05 is arbitrary and has been discouraged. The use of p-values in science is also controversial, because the correct interpretation of a p-value is

⁸⁷⁵ L Epstein, G King, 'The Rules of Inference' (2002) U Chi L Rev 1.

⁸⁷⁶ *ibid*, 37.

⁸⁷⁷ RL Wasserstein and NA Lazar 'The ASA's Statement on P-values: Context, Process and Purpose' (American Statistical Association 2016).

⁸⁷⁸ *ibid*.

complex.⁸⁷⁹ As a result, the rejection of null hypotheses in this study, owing to different levels of p-values, should not be read as being strong evidence that those particular variables influenced judicial decision-making. It should instead be read as a descriptive finding that in this particular sample and in this particular study there appears to be a statistical relationship between some variables and outcomes.

The problems relating to the use of p-values has been mitigated in this study. This study has utilised other measures of fit, such as classification tables and confidence intervals. Moreover, in explorative studies such as the present, p-values are a useful way to describe and to record the presence of patterns in data. Whilst this study has taken account of the ASA's statement, and the controversy relating to the use of p-values, there has been little mention of the statement within the ELS community. For instance, at the time of writing, there has been no discussion of the statement in the *Journal of Empirical Legal Studies*, and a search of legal databases found no mention of it. That particular journal is aware of the arbitrary character of statistical significance, however, with a 2013 foreword calling for authors to present effects with measures of statistical uncertainty.⁸⁸⁰ It stressed that 'conventional levels of statistical significance' are simply academic, and that descriptive analyses of law utilising statistics can also be useful. Thus, whilst there are inherent limitations within the methods used to model impartiality, there remains value in the observations of patterns within data. This gives rise to rich opportunities for research in the future.

A further reason why it may be said that impartiality has not been fully validly modelled, and so conclusions regarding the impartiality of the Court must be drawn cautiously, relates to the variables used. As was stated in Chapters 5 and 6, there are variables which could measure a concept such as impartiality which were not used in this study. There are several reasons why certain variables

⁸⁷⁹ See S Greenland, C Poole, 'Problems in Common Interpretations of Statistics in Scientific Articles, Expert Reports, and Testimony' (2011) (51) *Jurimetrics* 113.

⁸⁸⁰ DE Ho 'Foreword: Conference Bias' (2013) (10) 4 *JELS* 603.

were not used: difficulty in reliably capturing the data from the Court judgments; data simply not being available; and the constraints of time needed to collect data. In any study such as the present, a range of decisions need to be made as to which variables will be included in the study. It will always be possible to argue that alternative variables could have been used. This study has provided an initial range of variables, and has shown what variables, for the particular cases in this study, have an association with the outcome of appeals. As other variables could have been used, it is impossible to say that impartiality has been measured with precision. It will be the task of research in the future to determine whether any of these associations also exist in other samples of cases.

Within the variables which were utilised in this study, it may have been possible to have coded them differently. This further reduces how objective the variables are, and so how closely impartiality has been measured. As was shown in Chapter 6, some of the variables did require an element of subjective decision-making in their design and the collection. This was, however, for a minority of variables. Most of the variables were binary variables, which could be collected easily with the aid of the data collection template. If some variables were to be coded differently, it is possible that the level of association between that variable and the outcome variable could change. This could in turn have an impact on the conclusions of the study. Accordingly, the results of this study are best understood as an analysis of whether the particular range of variables used, including how they are coded, which may not be agreeable to every observer, are statistically associated with the outcome of appeals.

The 'Did Error Occur?' variable is an important variable in this study. It was shown to be the independent variable which was most strongly associated with the outcome of appeals against conviction. This variable has been presented as a measurement of the role of law, or a measure of the 'legal model'. Thus, the variable is interpreted as being a counterweight to factual and demographic variables, and an indicator of impartial decision-making. It may be challenged how well this variable does measure the Court's application of law. The finding

of a relationship between the 'Did Error Occur?' variable and the outcome of appeals is similar to previous findings in studies of the Court. Previous studies have found that appeals which cannot point to an error of law are unlikely to be successful, and appeals which can point to an error are the most likely to be successful. Accordingly, it cannot be surprising that there was a strong association found in this study, and that the addition of this variable improved the accuracy of the models.

Whilst previous studies have suggested that findings such as this is evidence of the allegedly restrictive approach of the Court, this study has contested this conclusion. Instead, it has been argued that the finding that the presence of legal errors are associated with successful appeals indicates that judges follow the law and legal rules, including the aims of the 'unsafety test'. This is one interpretation of the 'Did Error Occur?' variable, but further research could be considered to explore this further. Accordingly, this variable could be considered a relatively modest measure of the law governing the case.

It is argued, therefore, that the relationship between the variables in this study and the principle of impartiality, could be considered tenuous. This is owing to inherent limitations within the methods used, and limitations within the variables and their interpretations. This has led to the conclusion that taken as a whole there is insufficient evidence to doubt the null hypothesis that the Court determines appeals in an impartial manner. The relationships which have been found within the data are indicative of patterns, but cannot defeat the null hypothesis. It may be argued that this study is closer to an analysis of whether a range of variables are predictors of the outcome of appeals, rather than an analysis of the impartiality of the Court. As was stated at the outset of this thesis, this is as expected and this was the core aim of the study. This is an initial analysis of decision-making in the Court of Appeal, and as such it would be unlikely that it would be possible to draw strong conclusions from it.

Accordingly, the core conclusions which can be drawn from this study are that there is evidence of emerging patterns in the data, but this is insufficient to cause a reasonable observer to doubt the impartiality of the Court until further studies are undertaken to confirm this.

8.5 The ‘unsafety test’ and theories of judicial decision-making

In this thesis, a number of legal theories and theorists of judicial decision-making have been referred to. Three of these theories, by Lord Bingham, Hart, and the Legal Realists, accepted that judges exercise a discretion when they decide cases. Lord Bingham argued that when the law does not determine the outcome of a case, judges exercise discretion if they have to decide what is fair and just to do in the particular case.⁸⁸¹ Hart argued that when judges have to determine the meaning of a statute to determine a hard case, occasionally the meaning will not be clear. Due to the ‘open texture’ of language, judges may have a discretion open to them when deciding what the outcome of litigation should be.⁸⁸² The Legal Realists argued that law is at least locally indeterminate, in such cases the law will not fully determine the outcome of litigation.⁸⁸³ That being the case, the Legal Realists sought to discover what did determine litigation if it was not the law. In contrast, Dworkin argued that judges do not exercise discretion in the same way as these writers argued. Instead, they must discover which litigant has a right to win, by deciding which outcome would best fit and justify the existing political mortality and law.⁸⁸⁴

The question which arises, therefore, is whether any of these theories are most accurately reflected in the decision-making of the Court of Appeal for these appeals, and how does the Court’s decision-making accord with legal theory? It has been argued throughout this thesis that the ‘unsafety test’ invites the judges to exercise their discretion, but this is controlled by the law. It is more likely,

⁸⁸¹ T Bingham, 2006 British Academy Maccabean Lecture, available from www.law.cf.ac.uk/newsandevents/transcripts/271005.pdf <accessed 07 August 2016>.

⁸⁸² HLA Hart, *The Concept of Law*, (3rd ed, Oxford University Press 2012). Chapter 7.

⁸⁸³ B Leiter, *Naturalizing Jurisprudence: Essays on American Legal Realism and Naturalism in Legal Philosophy*, (Oxford University Press 2007) 41

⁸⁸⁴ R Dworkin, *Law’s Empire* (Harvard University Press 1986).

therefore, that decision-making was accord with the theories of Bingham, Hart, or the Realists rather than Dworkin. It is submitted that there is relatively little dispute between Hart and Lord Bingham, and the Legal Realists. All agreed that the law had indeterminate edges in which judges must exercise discretion. Hart argued that owing to the nature of ordinary language there would be circumstances in which the law would prove indeterminate. This part of the ‘open-texture’ of language.⁸⁸⁵ ‘Open-textured’ appears a particularly appropriate term for the ‘unsafety test’ given its fact sensitivity and the judges’ exercise of judgment.

Lord Bingham’s description of judicial decision-making is well reflected in the decisions of the Court. In many ways this would be expected, since Lord Bingham gave a number of important judgments explaining the ‘unsafety test’.⁸⁸⁶ He argued that a judges have no discretion when deciding issues of law, but that ‘it is only when he reaches the stage of asking himself what is the fair and just thing to do or order in the instance case that he embarks on the exercise of a discretion’.⁸⁸⁷ As discussed in Chapter 4, some kinds of appeals are determined after answering a question of law, such as whether the trial should have been stayed. In those which are not, the judges in the Court of Appeal exercise a degree of discretion, because deciding what is ‘fair and just’ to do in the appeals is what is required by the ‘unsafety test’. This also accords with Dennis’s opinion that the core of the ‘unsafety test’ is a question for the judges whether they are satisfied in the factual accuracy and moral authority of convictions, and that the conviction was achieved in accordance with the rule of law.⁸⁸⁸

Although, as was explained in Chapter 3, it is difficult to provide a definition of Legal Realism, some Realists appeared to suggest that the law was only a minor

⁸⁸⁵ See Hart (n 882 above).

⁸⁸⁶ For instance, *R v Graham* [1997] 1 Cr App R 302, and *R v Pendleton* [2001] UKHL 66.

⁸⁸⁷ T Bingham, *The Business of Judging: Selected Essays and Speeches* (Oxford University Press 2000) 36.

⁸⁸⁸ I Dennis, ‘Fair Trials and Safe Convictions’ [2003] CLP 211.

constraint upon judicial decision-making.⁸⁸⁹ This more radical 'rule sceptic' view of Legal Realism cannot be supported by this thesis because, as has been shown, it is the measurement of the law which is the strongest predictor of the outcome of appeals against conviction. However, as Leiter argued, most Realists did not deny that legal rules determine cases, and instead argued that the law was only 'locally indeterminate,' in particular at the level of appellate review.⁸⁹⁰ The 'received view' of Legal Realism, which holds that the Realists believed judges exercise unfettered discretion in all cases, is a 'Frankified', or Critical Legal Studies, version of Realism, and was not held by most Realists. The Realists held empirical goals to discover whether non-legal factors, such as judicial personality or politics filled some of the gap in what determines cases. Lord Bingham accepted that the more senior the court, the more influence political considerations can have.⁸⁹¹

Whilst it has been explained above that the results of this study must be read cautiously, it has been found that there appears to be a relationship in the data between judges and outcomes. This could be interpreted to suggest that the personalities of particular judges may have a role in appeals. However, even if the limitations of the study are overlooked, it is not possible to accept that judicial personality has a major role, because it was the legal variable which had the greatest association with outcomes. It is submitted, therefore, that a radical Legal Realism cannot be accepted, but a more nuanced version, embodied by Lord Bingham and Hart, is best reflected in the decision-making of the Court of Appeal. When the law governing the case shows that certain errors have occurred, this will usually render convictions unsafe, the judges have no discretion and will allow appeals.

⁸⁸⁹ Jerome Frank's *Law and the Modern Mind* (Transaction Publishers 2009) could be seen as the strongest statement of 'rule scepticism'.

⁸⁹⁰ See Leiter (n 883). See also Introduction by Schauer in KN Llewellyn, *The Theory of Rules* (edited and compiled by F Schauer) (University of Chicago Press 2011), suggesting Llewellyn was not a rule sceptic.

⁸⁹¹ Bingham (n 887) 28.

8.6 Outcome of null hypothesis tests of judicial impartiality

The question addressed in Chapter 1 of this thesis was whether the Court appeared to have determined appeals against conviction in an impartial manner. This has was presented in the form of the following null hypothesis:

H^0 = The Court of Appeal appeared to have determined appeals in an impartial manner.

The alternative hypothesis was:

H^1 = The Court of Appeal appeared to lack impartiality.

The proposition that the Court was impartial was chosen as the null hypothesis in this study because there is a lack of previous empirical evidence on the Court to suggest any other *a priori* hypothesis would be suitable. These hypotheses were tested by conducting a large number of other null hypothesis tests on the independent variables. Each factual and demographic variable had its own associated null hypotheses that it was not a statistically significant predictor of outcomes in the Court. To be able to justifiably reject H^0 , and make a declaration that H^1 is more likely, extraordinary evidence would be required. It would need to be shown that a well informed observer would conclude that the Court was biased or lacking in impartiality.⁸⁹² As was explained above, this level of evidence is not found in this thesis, and could only be sustained following replication and repeated testing. Accordingly, there is insufficient evidence to cause a reasonable observer to justifiably conclude that the Court lacked impartiality. The patterns which were found in the data are insufficient to be able to reject the null hypothesis, but should be a foundation for further research and analysis.

⁸⁹² As per *Porter v Magill* [2002] 2 AC 357.

It was hypothesised that the 'Did Error Occur?' variable would have a strong relationship with successful appeals, if it was found that an error occurred. This was found to be the case. As explained above, this is not a surprising outcome, because it echoes previous studies of the Court. However, it has been argued that the 'Did Error Occur?' variable encapsulates the 'legal model' of judicial decision-making. Thus, similar to the findings of Sisk, Heise and Morriss in their study,⁸⁹³ 'the law remains the alpha and omega of judicial decision-making.'⁸⁹⁴ Similar to Cross, 'legal rules of procedure matter greatly in determining outcomes'.⁸⁹⁵ Thus, a behavioural, attitudinal, or a radical Legal Realist perspective, of unfettered discretion or ideological decision-making cannot be supported by the results of this thesis.

Despite most of the variables having little association with successful appeals, some variables were statistically significant predictors. One potential explanation of this is that legal rules determine most appeals, but that there is a gap – the law does 'run out'. Clearly, fresh evidence appeals, which often do not raise procedural irregularities, account for part of this gap. In fresh evidence appeals the outcome is largely based upon whether the judges think the fresh evidence could have led to a different outcome. Appeals which raise procedural irregularities usually require the judges to determine whether the error makes the conviction unsafe. It is in this area, the point at which judges have decided that an error did occur and need to decide whether that makes the conviction unsafe, that judges have to exercise their judgement relating to the facts of the case. Research in the future could be conducted to further scrutinise this particular gap and exercise of judgment. It may be in these cases that the personality of the judge, or behavioural and attitudinal factors, could play a role.

⁸⁹³ G Sisk, M Heise and P. Morriss, 'Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning' (1998) 37 NYU L Rev 1377.

⁸⁹⁴ *ibid*, 1500.

⁸⁹⁵ FB Cross, *Decision Making in the US Courts of Appeals* (Stanford University Press 2007) 67.

8.7 Conclusion

This chapter has sought to explain what the results of this thesis mean about decision-making in the Court of Appeal. It has been shown that appeals against murder and rape convictions raise broadly similar issues as found in previous studies of the Court of Appeal. Appeals against murder and rape convictions were dominated by issues relating to the summing up, and claims that the judge incorrectly exercised his discretion to admit or exclude evidence. Lurking doubt appeals were also relatively rare, and were shown to be unlikely to be successful. The major difference in grounds of appeal raised was that fresh evidence was more frequently raised, and more frequently successful, in the murder and rape appeals than was seen in earlier studies. Given the significance which is placed on the finding that fresh evidence is usually rare, this has some importance. Where fresh evidence was raised in the murder and rape appeals, it was successful almost as frequently as it was unsuccessful. This means that it is not possible to support the conclusion that the Court was restrictive, across the cases in this sample, to fresh evidence.

The innovation in this thesis is that as well as collecting grounds of appeal, which, as was noted, is not necessarily an easy or a neutral task, other factual, demographic, and legal variables were collected. This allowed for binary logistic regression analysis of whether the variables collected were predictors of the outcomes of appeals. It was found that the law is the greatest predictor of success. However, factual and demographic factors, in particular the judges hearing the case and bad character, were shown also to be predictors of appeals. These results were analysed closely in order to give scrutiny to the findings. It has been submitted that there is insufficient evidence from this study to justifiably challenge the impartiality of the Court. A number of patterns have emerged from the data, which should give cause for research in the future.

Chapter 9

Thesis Conclusions

Introduction

The research question raised in Chapter 1 of this thesis was whether the Court of Appeal determined appeals against conviction in murder and rape appeals in an impartial manner. It was sought to explore this question by collecting data in relation to a range of variables. This is a critical question, because impartiality is the minimum standard expected of the courts if they are to have legitimacy. In Chapter 2 it was explained that this is an objective test. From *Porter v Magill*,⁸⁹⁶ the test is whether 'a fair-minded and informed observer, having considered the facts, would conclude that there was a real possibility that the tribunal was biased'.⁸⁹⁷ Prior to this study, there was little direct quantitative empirical analysis of this question in relation to the senior British judiciary. This thesis has examined the literature on the England and Wales Court of Appeal (Criminal Division), scrutinised the principle of impartiality, and explained how the analysis of impartiality was conducted.

It has been concluded that there is insufficient evidence for a fair-minded and informed observer to think that the Court was biased or lacking in impartiality. The study has, however, uncovered interesting evidence of patterns and relationships between certain variables and the outcome of appeals. This is not enough to justifiably begin to doubt the impartiality of the Court, but may lead to a basis for further research. In this chapter there is a summary of the original contributions made by this thesis, and an overview of the research findings. The relationship between this thesis and previous research is discussed. It is explained what can be concluded about the decision-making in the Court in light of the newly extended corpus of research. Avenues for future research, and the strengths and limitations of this study, are also discussed.

⁸⁹⁶ [2002] 2 AC 357.

⁸⁹⁷ *ibid.*

9.1 Summary of original contributions

This thesis has taken what has previously been written about the Court of Appeal, and asked an alternative question regarding its decision-making. Whilst previous studies were concerned with how well the Court performs in correcting miscarriages of justice,⁸⁹⁸ this thesis was concerned with the impartiality of the Court's decision-making. This thesis has explained in a comprehensive manner how the question of the Court's impartiality was measured and answered. It has been explained that the concept of impartiality has been incompletely measured, and so it is not possible to draw strong conclusions regarding impartiality. Accordingly, this study complements, but has analysed the Court of Appeal from a very differing perspective in respect to previous studies. This study has been designed to study the Court of Appeal from an Empirical Legal Studies (ELS) perspective, with the aim of making a methodological contribution. If looking for an objectivist, quantitative, Empirical Legal Studies analysis of the England and Wales Court of Appeal, this thesis should be the starting-point.

There were two elements to the empirical analysis of the Court of Appeal. Firstly, previous research relating to the grounds of appeal raised has been replicated. Secondly, binary logistic regression analysis of the factors which predict successful appeals in appeals against conviction in murder and rape appeals was conducted. The first aspect of this study broadly supported the findings of previous research. It has been shown that in murder and rape appeals, procedural irregularity grounds are the most commonly argued. This supports previous research which has highlighted the grounds of appeal raised.⁸⁹⁹ While this element of the research generally supported the findings of earlier studies, it was not wholly supportive. In particular, it was shown that fresh evidence appeals were raised more frequently in murder and rape appeals than was seen in previous studies. This is an important finding because the success rate of fresh evidence has some importance in earlier studies regarding the apparent

⁸⁹⁸ See S Roberts, *The Decision Making Process of Appeals Against Conviction in the England and Wales Court of Appeal (Criminal Division)* (DPhil thesis, London School of Economics and Political Science, 2009); SJ Heaton, *A Critical Evaluation of Using Innocence as a Criterion in the Post-Conviction Process* (DPhil Thesis, University of East Anglia, 2013).

⁸⁹⁹ *ibid.*

approach of the Court. As such, the results of this study did not necessarily support the claim of a restrictive approach because one of the factors seen to be pointing towards a restrictive approach in previous studies, namely a restrictive approach to fresh evidence appeals, was not seen in this study.

The second element of this thesis was to address the question of whether the Court of Appeal appeared to have decided the appeals against conviction in an impartial manner. It has been explained how that question was addressed in a methodological manner. In order to answer this question, a dataset of 472 full appeals against conviction, for the offences of murder and rape, was created. The creation of this dataset, in particular the methodologies behind its creation, represents a further original contribution to knowledge specifically from a methodological standpoint. It has been explained how a normative concept – impartiality – needs to be made concrete so that data can be collected. The key to making a normative concept concrete, and so susceptible to analysis, is to identify the ‘observable implications’ of the concept.

Based partly upon previous ELS research, and partly upon experience of reading Court of Appeal judgments, a range of variables were collected. ELS research has sought to test a range of models of judicial decision-making, including the legal, behavioural, attitudinal, and institutional models. Variables from these models have been drawn upon in this thesis. In combination, these variables sought to capture the principle of impartiality. The ‘legal model’ is most closely based upon the idea that judges decide appeals impartially. The ‘Did Error Occur?’ variable sought to capture elements of this model, and acted as a measure of whether the Court did determine appeals in an impartial manner. Variables drawn from the behavioural and attitudinal models were those which would not appear consistent with impartial decision-making.

Using the observed outcome of the appeal (successful or unsuccessful) as the dependent variable, the binary logistic regression analyses undertaken provides

the odds that a particular case will fall into a particular category (successful or unsuccessful), given the presence of particular independent variables. When this analysis was conducted, the majority of the independent variables were weak predictors of successful appeals. The lack of a strong relationship between behavioural variables is a potential indicator of impartial decision-making. The view that the Court could generally be considered to have appeared impartial was strengthened when the measurement of the law was included in the analysis. This is an important variable, because it acted as a counterweight variable for the factual and demographic variables whose statistical significance points to a departure from impartiality.

Judges are supposed to apply the law, and so any analysis which did not offer some role for the law would be much weaker. It was found that the law governing the case, measured by the answer to the question: in the opinion of the judges, 'Did an Error Occur?' was a strong predictor variable of case outcomes. The law determined a large proportion of appeals against conviction, leaving factual and demographic details with reduced predictive power with the models developed in the study. Moreover, the finding that factual and demographic variables had a limited role when fresh evidence appeals were considered in isolation, and when CCRC appeals were removed from the dataset.

Although it may be concluded that the Court did appear impartial, some variables drawn from the behavioural model showed an association with the outcome of appeals. These included the judges an appellant had, and whether the appellant was stated to have had bad character. The question of objective impartiality asks whether reasonable observers of the Court, after being informed of the facts, can be confident that the Court appeared to have decided cases in an impartial manner. It was ultimately suggested that, despite the finding of statistical significance of behavioural variables, that is insufficient for an objective and informed observer to question the appearance of the impartiality of the Court.

9.2 Decision-making in the England and Wales Court of Appeal (Criminal Division)

Since 1995, the sole test for the Court of Appeal when seeking to discharge its duty of doing justice is whether the conviction is 'unsafe'.⁹⁰⁰ This thesis has sought to investigate one specific aspect of the Court's decision-making: whether it acted in an impartial manner. Since the enactment of the unsafety test, there have been two empirical analyses of the approach of the Court: by Roberts,⁹⁰¹ and Heaton.⁹⁰² Prior to this was Malleson's study on behalf of the *Royal Commission on Criminal Justice*.⁹⁰³ Critical analysis of the Court's decision-making has suggested that the Court is too slow to overturn jury verdicts when nothing went wrong at trial; that it is unduly deferential to the verdict of the jury and the principle of finality; and that it is fearful of opening the floodgates.⁹⁰⁴ These are extremely serious allegations to make against the Court, and is suggestive of favouring legal principles, such as finality, over doing substantive justice. As has been explained, analysis of the approach of the Court has been explicitly eschewed in this thesis as it does not appear to be a stable standard against which to review the Court's decision-making.

The operation of the 'unsafety test' is based upon whether the particular judges hearing the appeal, at the particular time that the appeal is heard, are sure of the factual accuracy of the conviction; its moral authority; and that the conviction is grounded in the rule of law.⁹⁰⁵ It is only if the Court can answer these questions affirmatively that convictions can be safe. If the judges have anything more than a lurking doubt about guilt, up to and including being sure that the appellant is *not*

⁹⁰⁰ Criminal Appeal Act 1968, s 2. (As amended by Criminal Appeal Act 1995).

⁹⁰¹ S Roberts, *The Decision Making Process of Appeals Against Conviction in the England and Wales Court of Appeal (Criminal Division)* (DPhil thesis, London School of Economics and Political Science, 2009); S Roberts, 'The Royal Commission on Criminal Justice and Factual Innocence: Remediating Wrongful Convictions in the Court of Appeal' (2004) (1)(2) JJ 86.

⁹⁰² SJ Heaton, *A Critical Evaluation of Using Innocence as a Criterion in the Post-Conviction Process* (DPhil Thesis, University of East Anglia, 2013).

⁹⁰³ K Malleson, *The Royal Commission on Criminal Justice Review of the Appeal Process*, Research Study No 17 (HMSO 1993).

⁹⁰⁴ See S Roberts and L Weathered, 'Assisting the Factually Innocent: The Contradictions and Compatibility of Innocence Projects and the Criminal Cases Review Commission' (2009) 29 OJLS 43

⁹⁰⁵ See I Dennis 'Fair Trials and Safe Convictions' (2003) CLP 211.

guilty, the conviction will be unsafe.⁹⁰⁶ The power to determine appeals against conviction is vested in judges; that is their role. They are required to determine appeals based on all the circumstances as they see them, in an impartial manner.⁹⁰⁷ This thesis is the first attempt to model the behaviour of the Court using advanced inferential statistics in order to explore its decision-making in murder and rape appeals against conviction.

The results of this thesis suggests that most factual and demographic variables have no, or little, association with the judges' decisions of where justice lies. Some evidence has been uncovered which suggests that certain judges may be more likely to find convictions to be unsafe, and others may be more likely to find convictions to be safe. However, the regression analysis conducted is correlative and does not imply causation. The emergence of patterns found within the data give grounds for further analysis of the issues raised.

9.3 Strengths and limitations of study

The work conducted in this thesis has a number of strengths. The dataset developed for the purposes of the analysis was relatively large, and within the range of dataset sizes used in previous ELS research. Due to the way in which the data was collected, utilising the template with multiple fields discussed in Chapter 6, a wide variety of factual, demographic and legal variables could be collected. By utilising the purposive selection procedure it was possible to determine at an early stage which variables showed some predictive strength for inclusion in later models. The collection of the variables was driven by previous ELS research into the decision-making of judges, coming primarily from the United States. This embedded the research study within the ELS movement, utilising their methods and learning from their mistakes. The research was not, therefore, conducted in a vacuum, but has been conducted in line with ELS research which has a long history. This is important because utilising these

⁹⁰⁶ See JC Smith, 'The Criminal Appeal Act 1995: Part 1: Appeals Against Conviction' (1995) Crim LR 920.

⁹⁰⁷ See Article 6 European Convention on Human Rights.

methods to study the England and Wales Court of Appeal was an innovative step, and fairly alien to the British legal academy. By conducting the research within a movement with heritage, concerns regarding the feasibility of the study can be dispelled.

The primary limitations of the study are inherent within the chosen methodology. There were no circumstances in which to replicate experimental conditions, and so the study had to be conducted observationally. This resulted in a sense of artificiality, as only part of the decision-making process (that pertaining to the factual, demographic, and legal variables used in the analysis) could be observed. This imposes difficulties as there could always be other variables which would explain Court decisions which were not captured in the data collection process. Indeed, the pseudo R^2 statistics presented in Chapter 7 suggests that not all the variation in Court outcomes has been explained by the models, and so there may be other variables which may explain more variation in the outcomes.

This study has been presented as an exploration of the Court's impartiality. However, it has been questioned how successfully impartiality has been measured. As discussed in Chapter 8, impartiality has not been completely measured. There are other variables which could have been utilised to provide a fuller approximation of the concept. Furthermore, coding decisions were made with which not all observers will agree. This means that the results regarding the concept of impartiality must be interpreted cautiously. The study modelled a range of variables in order to explore the relationship between them and the outcome of appeals. These variables are related to but do not fully encapsulate the principle of impartiality. Within these variables there is evidence of patterns within the data. Some of these patterns are indicative of impartial decision-making, and some patterns may be suggestive otherwise. However, as impartiality is incompletely measured, it cannot be said that this resolves the question of the Court's impartiality. Instead, patterns have emerged between

certain variables and outcomes, which will require further development in the future.

9.4 Potential avenues for future research

This thesis has created a number of potential avenues for future research. Initially, it could be sought to replicate this study, or it could be sought to utilise the methods of this thesis to study the approach of the Court. This thesis has highlighted a number of factual, demographic, and legal factors which were associated with the outcome of appeals. In order to validate the results of this thesis, the analysis will need to be replicated. The analysis could be replicated by analysing the same variables across different offences or periods of time. Future research could also seek to improve upon some of the variables which were collected. There were a number of important variables which could not be collected in this thesis, this is a limitation of the study and could be sought to be rectified. For instance, ethnicity of the parties to appeals was not collectable, and data relating to the trial judge and sentencing was often missing. Techniques to remediate these difficulties could be further considered, for instance by systematically interviewing judges or other professionals working in the Court.

Moreover, a fuller account of the role of law in judicial decision-making could be considered. The legal variable used in this thesis could be improved by, for instance, considering more closely the role of precedent in forming the rules which govern the Court. If the legal variable could be enhanced, and further behavioural variables could be collected, then impartiality may be more closely measured. This would mean that impartiality would be measured with more validity, and so stronger conclusions regarding impartiality could be drawn.

9.5 Conclusion

This study sought to address the impartiality of the decision-making of the England and Wales Court of Appeal (Criminal Division) by observing patterns emerging from data collected from Court judgments. By conducting research within an Empirical Legal Studies framework, this study has embarked upon an exploration of what kinds of factors drive judicial decision-making. It has been shown that the strongest relationships were between outcomes and variables which suggested impartial decision-making. The presence of other relationships in the data warrants further analysis, and it is intended that this study will contribute to the development of this conversation in Britain.

APPENDICES

Appendix A

List of Variables

Factual and demographic variables

- The day of the week and month of the year of the appeal
- Whether the appellant is a child
- Whether appellant was of previous good character
- Whether appellant was of previous bad character
- The deceased / complainant age range
- The ranks of the judges
- Complainant / deceased gender
- Whether the appellant was convicted of other offences
- Whether appellant convicted of multiple counts of rape or murder
- Nature of the defence: whether denial of *actus reus* or *mens rea* or both
- Whether drink or drugs were involved
- Whether verdict *ex tempore*
- Whether only the appellant had a female counsel
- Whether only the Crown had female counsel
- Whether a female judge present on the bench
- Whether offence an historical offence
- The individual judges who heard at least 20 appeals
- Whether appeal was a joined appeal
- Whether deceased / complainant were known to the appellant
- Whether only the appellant was represented by a QC
- Whether only the Crown was represented by a QC
- Sentence severity
- Whether trial judge had sat in the Court of Appeal
- White British Appellant

Institutional variables

- CCRC reference
- Unanimous jury verdict
- How leave was granted

Legal variables

- Did error occur?
- Number of Cases cited
- Offence

Appendix B

List of Sampled Appeals

2006 EWCA Crim [...]

67	657	56	179	1359	8	14	454
193	231	623	408	165	139	3427	195
862	441	556	542	417	756	868	691
786	809	1220	627	268	905	706	866
946	793	962	819	1275	821	853	1294
1325	1503	1509	1233	1292	955	1267	1973
1971	1155	1824	1419	1941	1622	1652	1371
1645	1353	1498	2265	1705	1720	2029	1860
1721	2052	1703	2308	1976	1928	1887	1980
1891	1395	1843	1901	2120	2150	1884	2351
2325	2101	2441	2442	2317	2545	3364	3044
2945	3038	2636	2800	2752	2792	2743	2898
2625	2984	2983	2988	2914	2919	2749	3326
3227	3222	3092	3336	3182	3120	3249	3378
3360	3388	3187	3347	3381			

2007 EWCA Crim [...]

23	1	30	57	185	329	237	219
462	378	518	709	985	597	710	342
684	481	691	966	804	1122	764	1182
942	2605	1620	1779	1275	1184	1251	1250
1837	2064	1237	3201	1619	1669	1581	1494
1558	1665	1743	1698	1471	1756	1750	1709
1680	1556	1912	1699	1702	1899	1715	1943
2687	2083	3234	2018	2264	2109	2066	2236
2176	2283	2660	2551	2739	2633	2631	2974
2887	2741	2881	2581	2865	3198	3083	2722
3029	3473	3027	3226	2870	3437	3282	3047
3323	3132						

2008 EWCA Crim [...]

219	4	6	266	76	312	183	3305
587	460	866	559	593	544	434	603
688	754	529	585	678	939	968	817
1359	879	994	1033	923	1126	1092	971
973	1014	1329	1415	1111	1370	1156	2015
1222	1305	1342	1801	1851	1767	1792	1850
1923	2121	1739	3229	2552	2774	2874	3249
3018	2563	2564	2946	2788	3949	2787	2989
3147	3238	3059					

2009 EWCA Crim [...]

52	65	208	165	64	20	19	255
316	86	236	50	190	805	280	213
404	513	434	678	490	594	593	910
741	742	819	734	1187	1240	789	905
1370	1086	1057	1239	1208	1254	1130	1165
1211	930	1531	1372	1076	1308	1457	1371
1326	1617	1362	1425	1960	1600	1721	1718
1569	1829	1638	1739	1876	2137	2342	2345
2368	2520	2530	2425	2243	2294	2291	2457
2458	2552	2728	2553	2705	2688	2804	2909
2695	2668	2732	2781	2878	2698	2969	2697
2698							

2010 EWCA Crim [...]

7	3	4	119	203	312	2402	90
194	138	147	350	195	1228	226	445
385	525	416	450	909	496	516	777
547	865	755	1034	929	830	1251	2353
926	832	1380	1379	914	1176	2317	1230
1082	1150	1152	1213	1149	1417	1325	1269
1589	1461	1579	1683	1595	1768	1578	1620
1623	1622	1692	1836	1859	2259	2039	1818
2299	1691	1856	1924	1980	3271	2144	2101
2134	2349	2358	2072	2647	2269	2445	2981
2620	2438	3335	2553	2725	3339	2664	3096
2740	3067	2847	3117	2899	2949	2936	3089
2975	2972	2950	2974	2499			

Other cases:

2009 All ER 47

Appendix C

Purposive Selection Results

Joined Case

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Joined_Case	-0.526	0.354	2.215	1	0.137	0.591
	Constant	-0.860	0.107	64.423	1	0.000	0.423

Monday

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Mon	0.243	0.356	0.464	1	0.496	1.275
	Constant	-0.936	0.107	76.731	1	0.000	0.392

Tuesday

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Tues	-0.141	0.248	0.320	1	0.572	0.869
	Constant	-0.884	0.115	59.171	1	0.000	0.413

Wednesday

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Wed	-0.429	0.258	2.765	1	0.096	0.651
	Constant	-0.824	0.114	52.366	1	0.000	0.439

Thursday

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Thur	0.099	0.241	0.169	1	0.681	1.104
	Constant	-0.937	0.116	65.041	1	0.000	0.392

Friday

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Fri	0.314	0.232	1.830	1	0.176	1.369
	Constant	-0.994	0.119	69.893	1	0.000	0.370

January

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Jan	0.243	0.356	0.464	1	0.496	1.275
	Constant	-0.936	0.107	76.731	1	0.000	0.392

February

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Feb	0.425	0.324	1.722	1	0.189	1.530
	Constant	-0.959	0.108	78.530	1	0.000	0.383

March

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	March	-0.140	0.352	0.158	1	0.691	0.869
	Constant	-0.902	0.107	71.108	1	0.000	0.406

April

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Apr	-0.690	0.503	1.885	1	0.170	0.502
	Constant	-0.879	0.104	70.914	1	0.000	0.415

May

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	May	0.105	0.303	0.120	1	0.729	1.111
	Constant	-0.928	0.109	72.243	1	0.000	0.395

June

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	June	-0.202	0.364	0.307	1	0.579	0.817
	Constant	-0.897	0.107	70.826	1	0.000	0.408

July

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	July	-0.027	0.299	0.008	1	0.928	0.973
	Constant	-0.911	0.109	69.278	1	0.000	0.402

August

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Aug	-0.887	1.085	0.669	1	0.413	0.412
	Constant	-0.904	0.102	77.998	1	0.000	0.405

September

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Sept	-0.601	0.788	0.582	1	0.446	0.548
	Constant	-0.903	0.103	77.104	1	0.000	0.405

October

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Oct	0.059	0.375	0.025	1	0.874	1.061
	Constant	-0.920	0.106	74.956	1	0.000	0.399

November

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Nov	0.322	0.338	0.908	1	0.341	1.379
	Constant	-0.946	0.108	77.326	1	0.000	0.388

December

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Dec	-0.172	0.351	0.240	1	0.624	0.842
	Constant	-0.898	0.107	70.520	1	0.000	0.407

Ex-tempore

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Ex_tempore	0.055	0.214	0.065	1	0.799	1.056
	Constant	-0.950	0.174	29.972	1	0.000	0.387

QC Appellant Only

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	QC_APP_Only	0.056	0.317	0.032	1	0.859	1.058
	Constant	-0.921	0.108	72.254	1	0.000	0.398

QC Crown Only

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	QC_Crown_Only	-0.505	0.435	1.345	1	0.246	0.604
	Constant	-0.881	0.105	70.298	1	0.000	0.414

F Counsel Appellant

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	F_Counsel_App	0.043	0.249	0.029	1	0.864	1.044
	Constant	-0.924	0.115	64.755	1	0.000	0.397

F Counsel Crown

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a F_Counsel_Crown	0.214	0.230	0.863	1	0.353	1.238
Constant	-0.970	0.119	66.252	1	0.000	0.379

F Judge Present

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a F_Judge_Present	0.049	0.225	0.047	1	0.828	1.050
Constant	-0.929	0.121	59.353	1	0.000	0.395

LCJ Present

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a LCJ_Present	0.124	0.415	0.089	1	0.765	1.132
Constant	-0.923	0.105	76.749	1	0.000	0.397

VP Present

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a VP_Present	-0.163	0.382	0.182	1	0.670	0.850
Constant	-0.902	0.106	72.307	1	0.000	0.406

President Present

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a President_Present	-0.440	0.513	0.736	1	0.391	0.644
Constant	-0.895	0.104	73.836	1	0.000	0.409

Circuit Judge Present

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a Circuit_Judge_Present	-0.211	0.262	0.648	1	0.421	0.810
Constant	-0.874	0.113	59.821	1	0.000	0.417

Recorder Present

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a Recorder_Present	-0.209	0.317	0.435	1	0.509	0.811
Constant	-0.889	0.108	67.255	1	0.000	0.411

Retired Present

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Retired_Present	0.411	0.308	1.784	1	0.182	1.509
	Constant	-0.963	0.109	77.934	1	0.000	0.382

Trial Judge CA

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Trial_J_CA	-0.404	0.273	2.184	1	0.139	0.668
	Constant	-0.862	0.128	45.447	1	0.000	0.422

Judge 1

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Dobbs	-0.495	0.511	0.939	1	0.333	0.610
	Constant	-0.891	0.104	73.250	1	0.000	0.410

Judge 2

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Gage	0.962	0.459	4.382	1	0.036	2.616
	Constant	-0.962	0.105	83.628	1	0.000	0.382

Judge 3

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Hallett	-1.088	0.545	3.990	1	0.046	0.337
	Constant	-0.858	0.104	67.748	1	0.000	0.424

Judge 4

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Hooper	0.970	0.324	8.957	1	0.003	2.638
	Constant	-1.016	0.109	86.470	1	0.000	0.362

Judge 5

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Hughes	-0.115	0.403	0.081	1	0.775	0.891
	Constant	-0.907	0.106	73.790	1	0.000	0.404

Judge 6

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Judge	0.113	0.363	0.098	1	0.755	1.120
	Constant	-0.924	0.107	75.247	1	0.000	0.397

Judge 7

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Latham	0.071	0.499	0.020	1	0.888	1.073
	Constant	-0.918	0.104	77.658	1	0.000	0.399

Judge 8

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Moses	0.292	0.405	0.520	1	0.471	1.339
	Constant	-0.934	0.106	78.232	1	0.000	0.393

Judge 9

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Openshaw	-0.133	0.486	0.075	1	0.785	0.876
	Constant	-0.909	0.104	75.884	1	0.000	0.403

Judge 10

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Rafferty	-0.250	0.480	0.272	1	0.602	0.778
	Constant	-0.902	0.104	74.705	1	0.000	0.406

Judge 11

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Thomas	-0.547	0.508	1.157	1	0.282	0.579
	Constant	-0.888	0.104	72.665	1	0.000	0.411

Judge 12

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Judge 12	0.665	0.329	4.080	1	0.043	1.944
	Constant	-1.234	0.237	27.124	1	0.000	0.291

Cases Cited

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a	Cases_Cited	-0.005	0.022	0.053	1	0.818	0.995
	Constant	-0.895	0.124	52.095	1	0.000	0.408

Offence (Rape)

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a Offence	-0.583	0.207	7.968	1	0.005	0.558
Constant	-0.635	0.138	21.103	1	0.000	0.530

Low Sentence

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a Sentence_Low	0.456	0.272	2.816	1	0.093	1.578
Constant	-1.110	0.124	79.724	1	0.000	0.330

Medium Sentence

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a Sentence_Med	-0.019	0.220	0.008	1	0.930	0.981
Constant	-1.012	0.156	42.025	1	0.000	0.364

High Sentence

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a Sentence_High	0.017	0.273	0.004	1	0.950	1.017
Constant	-1.025	0.124	68.783	1	0.000	0.359

Very High Sentence

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a Sentence_V_High	-0.833	0.425	3.844	1	0.050	0.435
Constant	-0.935	0.115	66.541	1	0.000	0.393

Good Character

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a App_GC_Stated	0.189	0.256	0.547	1	0.460	1.208
Constant	-0.951	0.114	69.834	1	0.000	0.386

Bad Character

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a Bad_Character	-0.652	0.213	9.392	1	0.002	0.521
Constant	-0.648	0.130	24.785	1	0.000	0.523

Denial of AR

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	DENIAL_OF_AR	-0.139	0.207	0.451	1	0.502	0.870
	Constant	-0.832	0.159	27.503	1	0.000	0.435

Denial of MR

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Denial_of_MR	0.306	0.210	2.107	1	0.147	1.357
	Constant	-1.026	0.130	62.780	1	0.000	0.358

CCRC

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	CCRC	1.128	0.321	12.335	1	0.000	3.090
	Constant	-1.037	0.110	88.966	1	0.000	0.354

Single Judge

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Single_Judge	-0.658	0.206	10.172	1	0.001	0.518
	Constant	-0.550	0.149	13.593	1	0.000	0.577

Historical Offence

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Historical_Offence	0.519	0.294	3.122	1	0.077	1.680
	Constant	-0.983	0.110	79.466	1	0.000	0.374

Appellant under 18

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Appellant_under_18	-0.009	0.010	0.842	1	0.359	0.991
	Constant	-0.923	0.102	81.233	1	0.000	0.397

Under 13

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	UNDER 13	0.402	0.234	2.951	1	0.086	1,494
	Constant	-1.013	0.119	72.946	1	0.000	0.363

Age 13_Under 16

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a Age 13_Under_16	0.239	0.384	0.387	1	0.534	1.270
Constant	-0.932	0.106	77.334	1	0.000	0.394

16 - 17

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a 16_17	0.042	0.391	0.012	1	0.914	1.043
Constant	-0.918	0.106	75.259	1	0.000	0.399

Known Victim

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a KNOWN	0.045	0.275	0.026	1	0.871	1.046
Constant	-0.952	0.251	14.386	1	0.000	0.386

Appendix D

Binary Logistic Regression Tables: Model of Legal Variable

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	27.720	21	0.148
	Block	27.720	21	0.148
	Model	27.720	21	0.148

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	485.810 ^a	0.063	0.090

Hosmer and Lemeshow Test

Step	Chi-square	Df	Sig.
1	16.741	8	0.033

Classification Table^a

Observed			Predicted		
			Did_Error_Occur		Percentage Correct
			Yes	No	
Step 1	Did_Error_Occur	Yes	10	115	8.0
		No	11	287	96.3
Overall Percentage					70.2

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Joined_Case	0.520	0.357	2.122	1	0.145	1.683
	Wed	-0.075	0.286	0.068	1	0.794	0.928
	Fri	0.327	0.279	1.371	1	0.242	1.386
	Feb	-0.298	0.431	0.479	1	0.489	0.742
	April	-0.304	0.481	0.399	1	0.528	0.738
	QC_Crown_Only	-0.599	0.465	1.658	1	0.198	0.549
	Retired_Present	0.135	0.350	0.150	1	0.699	1.145
	Judge 1	0.454	0.521	0.759	1	0.384	1.575
	Judge 2	-0.510	0.496	1.056	1	0.304	0.601
	Judge 3	0.455	0.396	1.321	1	0.250	1.576
	Offence	0.310	0.303	1.050	1	0.306	1.364
	SENTENCE_LOW	0.332	0.296	1.260	1	0.262	1.394
	SENTENCE_V_HIGH	-0.482	0.405	1.418	1	0.234	0.617
	Bad_Character	0.457	0.246	3.452	1	0.063	1.580
	DENIAL_OF_MR	0.364	0.254	2.057	1	0.152	1.439
	CCRC	-0.095	0.462	0.042	1	0.837	0.909
	Single_Judge	-0.304	0.240	1.597	1	0.206	0.738
	Historical_Offence	0.804	0.358	5.028	1	0.025	2.233
	Under_13	0.255	0.396	0.414	1	0.520	1.290
	Adult	0.397	0.314	1.603	1	0.205	1.488
	UNANIMOUS	-0.357	0.275	1.676	1	0.195	0.700
	Constant	-0.393	1.542	0.065	1	0.799	0.675

Bibliography

Books

Ashworth, A., Redmayne, M., *The Criminal Process* (4th ed, Oxford University Press 2010)

Baker, BJ., *Textbook of Criminal Law* (3rd edition, Sweet & Maxwell 2012)

Banakar, R., Travers, M., (eds) *Theory and Method in Socio-Legal Research* (Hart Publishing 2005)

Banforado, A., *Unfair: The New Science of Criminal Injustice* (Crown Publishers 2015)

Baum, L., *Judges and Their Audiences: A Perspective on Judicial Behavior* (Princeton University Press 2008)

Bingham, T., *The Business of Judging: Selected Essays and Speeches* (Oxford University Press 2000)

Black, TR., *Understanding Social Science Research* (2nd edition, 2002, SAGE Publishing)

Blaikie, N., *Designing Social Research* (2nd edition, Polity Press 2010)

Cane, P., Kritzer, HM., (eds), *The Oxford Handbook of Empirical Legal Research*, (Oxford University Press 2010)

Clayton, C., Gillman, H., (Eds), *The Supreme Court in American Politics: New Institutional Interpretations* (University of Kansas Press 1999)

Cross, FB., *Decision Making in the US Courts of Appeals* (Stanford University Press 2007)

Cumming, G., *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis* (Routledge 2012)

Dagan, H., *Reconstructing American Legal Realism and Rethinking Private Law* (Oxford University Press 2013)

- Darbyshire, P., *Sitting in Judgment: The Working Lives of Judges*, (Hart Publishing 2011)
- Devlin, P., *The Judge*, (Oxford University Press 1979)
- Dicey, AV., *Introduction to the Study of the Law of the Constitution* (8th edition, Macmillan 1915)
- Duxbury, N., *Patterns of American Jurisprudence* (Clarendon 1995)
- Dworkin, R., *Law's Empire* (Harvard University Press 1986)
- , *Taking Rights Seriously* (Duckworth & Co, 1977)
- Epstein, L., Knight, J., *The Choices Justices Make* (CQ Press 1998)
- , Landes, WM., Posner, RA., *The Behaviour of Federal Judges: A Theoretical and Empirical Study of Rational Choice* (Harvard University Press, 2013)
- , and Martin, AD., *An Introduction to Empirical Legal Research* (Oxford University Press 2014)
- Field, A., *Discovering Statistics Using IBM SPSS Statistics* (4th ed, SAGE Publishing 2013)
- Frank, J., *Law and the Modern Mind* (Transaction Publishers 2009)
- Galligan, DJ., *Discretionary Powers: A Legal Study of Official Discretion* (Clarendon 1990)
- Garner, BA., *Garner's Dictionary of Legal Usage*, (3rd ed, Oxford University Press 2011)
- Gee, G., Hazell, R., Malleson, K., O'Brien, P., *The Politics of Judicial Independence in the UK's Changing Constitution* (Cambridge University Press 2015)
- Gelsthorpe, L., Padfield, N., (eds) *Exercising Discretion: Decision-Making in the Criminal Justice System and Beyond* (Willan Publishing 2003)
- Geyh, CG., *Courting the Peril: The Political Transformation of the American Judiciary* (Oxford University Press 2016)

- Gilligan, C., *In a Different Voice: Psychological Theory and Women's Development*, (Harvard University Press 1982)
- Griffiths, JAG., *The Politics of the Judiciary* (5th ed, Fontana 1997)
- Hammersley, M., *Methodology: Who Needs It?* (Sage, 2011)
- , *The Politics of Social Research* (Sage Publications 1995)
- Hart, HLA., *The Concept of Law*, (3rd ed, Oxford University Press 2012)
- Hosmer, DW, Lemeshow, S., Sturdivant, RX., *Applied Logistic Regression* (3rd ed, Wiley, 2013)
- Kahneman, D., *Thinking, Fast and Slow*, (Penguin 2011)
- Kennedy, D., *A Critique of Adjudication: fin de siècle* (Harvard University Press 1997)
- Klecka, WR., *Discriminant Analysis* (Sage Publications 1980)
- Klein, D, Mitchell, G., (Eds), *The Psychology of Judicial Decision-Making* (Oxford University Press 2010)
- Knight, M., *Criminal Appeals* (Stephen and Sons 1970)
- Leiter, B., *Naturalizing Jurisprudence: Essays on American Legal Realism and Naturalism in Legal Philosophy*, (Oxford University Press 2007)
- Llewellyn, KN., *The Common Law Tradition: Deciding Appeals* (Little, Brown & Co 1960)
- , *The Theory of Rules* (edited and compiled by F. Schauer) (University of Chicago Press 2011)
- Maveety, NL., (Ed), *The Pioneers of Judicial Behavior* (University of Michigan Press 2003)
- McBarnet, DJ., *Conviction: The Law, The State and the Construction of Justice* (Palgrave Macmillan, 1981)
- McConville, M., Bridges, L., (eds) *Criminal Justice in Crisis* (Edward Elgar 1994)
- , and Chui, WH., *Research Methods for Law* (Edinburgh University Press 2007)

- , Sanders, A., Leng, R., *The Case for the Prosecution* (Routledge, 1991)
- , *Standing Accused: The Organisation and Practices of Criminal Defence Lawyers in Britain* (Clarendon 1994)
- McGlynn, C., Hunter, R., Rackley, E., (eds) *Feminist Judgments From Theory to Practice* (Hart Publishing 2010)
- Mertz, E., Macaulay S., Mitchell, TW., (Eds) *The New Legal Realism: Translating Law and Society for Today's Legal Practice (Vol 1)* (Cambridge University Press 2016)
- Meyerson, D., *Understanding Jurisprudence* (Routledge 2007)
- Naughton M., (ed), *The Criminal Cases Review Commission: Hope for the Innocent?* (Palgrave Macmillan 2009)
- , *The Innocent and the Criminal Justice System: A Sociological Analysis of Miscarriages of Justice* (Palgrave Macmillan 2013)
- Neuman, WL., *Social Research Methods: Qualitative and Quantitative Approaches* (7th ed, Pearson Education 2014)
- Nobles, R., Schiff, D., *Understanding Miscarriages of Justice: The Law, The Media, and the Inevitability of Crisis* (Oxford University Press 2000)
- Norris, G., Qureshi, F., Howitt, D., Cramer, D., *Introduction to Statistics with SPSS for Social Science* (Pearson 2012)
- Packer, H., *The Limits of the Criminal Sanction* (Stanford University Press 1968)
- Pattenden, R., *English Criminal Appeals 1844-1994* (Clarendon Press 1996)
- Pinello, DR., *Gay Rights and American Law* (Cambridge University Press 2003)
- Raban, O., *Modern Legal Theory and Judicial Impartiality* (Glasshouse Press 2003)
- Rackley, E., *Women, Judging and the Judiciary: From Difference to Diversity* (Routledge 2013)
- Redmayne, M., *Character in the Criminal Trial* (Oxford University Press 2015)

- Riffe, D., Lacy, S., Fico, FG., *Analyzing Media Messages: Using Quantitative Content Analysis in Research* (2nd ed, Lawrence Erlbaum Associates 2005)
- Robertson, D., *Judicial Discretion in the House of Lords* (Clarendon Press 1998)
- Ross, RE., *The Court of Criminal Appeal* (Butterworth & Co 1911)
- Saldana, J., *Fundamentals of Qualitative Research*, (Oxford University Press 2011)
- Sanders, A., Young, R., Burton, M., *Criminal Justice*, (4th edition, Oxford University Press 2010)
- Schubert, GA., *Quantitative Analysis of Judicial Behavior* (Free Press 1959)
- , *The Judicial Mind: Attitudes and Ideologies of Supreme Court Justices 1946 – 1963* (Northwestern University Press 1965)
- Segal, JA., Spaeth, HJ., *The Supreme Court and the Attitudinal Model Revisited* (Cambridge University Press 2002)
- Sunstein, CR., *Why Societies Need Dissent* (Harvard University Press 2005)
- , D Schkade, LM Ellman, A Sawicki, *Are Judges Political? An Empirical Analysis of the Federal Judiciary* (Brookings 2006)
- Tamanaha, BZ., *Beyond the Formalist-Realist Divide: The Role of Politics in Judging* (Princeton University Press 2010)
- Trechsel, S., Summers, S., *Human Rights in Criminal Proceedings*, (Oxford University Press 2006)
- Twining, W., *Karl Llewellyn and the Realist Movement* (2nd ed, Cambridge University Press 2012)
- Tyler, TR., *Why People Obey the Law*, (2nd ed, Princeton University Press 2006)
- Walker, C., Starmer, K., (eds) *Justice in Error* (Blackstone 1993)

Journal articles

Adler, M., Simon, J., 'Stepwise Progression: The Past, Present, and Possible Future of Empirical Research on Law in the United States and United Kingdom' (2014) 41 J Law & Soc'y 173

Arvind, TT., L Stirton, 'Legal Ideology, Legal Doctrine and the UK's Top Judges' (2016) PL 418

Ashenfelter, O., Eisenberg, T., Schwab, S.J., 'Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes' ('995) 25 J of Legal Stud 257

Babiyak, MA., 'What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models' (2004) (66) Psychosomatic Medicine 411

Baxter, JS., Bain, SA, McAusland, JH., "'Are You Thick?": A Case Study of Confrontational Aspects Of A Police Interview With A Suspect' (2007) PJ 99

Blaxland, H., 'Sappers and Underminers: Fresh Evidence Revisited' (2017) Crim LR 537

—, and Wilcock, P., 'Fresh Evidence in Criminal Appeals – Pendleton Revisited' (2006) Arch News 4

Bloom, A., 'The "Post-Attitudinal Moment": Judicial Policy Making Through the Lens of New Institutionalism' (2001) 35 Law & Soc'y Rev 219

Bornstein, BH., Miller, MK., 'Does a Judge's Religion Influence Decision Making?' (2009) 45 Court Rev 112

Boyd, CL., Epstein, L., Martin, AD., 'Untangling the Causal Effects of Sex on Judging' (2010) (54)(2) American J of Pol Sci 389

—, 'In Defence of Empirical Legal Studies' (2015) 64 Buff L Rev 363

—, 'Representation on the Courts? The Effects of Trial Judges Sex and Race' (2016) 69 Polit Res Q 788

Bursac, Z., Gauss, CH., Williams, DK., Hosmer, DW., 'Purposeful Selection of Variables in Logistic Regression; (2008) (3) Source Code of Biology and Medicine 17

Cahill-O'Callaghan, RJ., 'Reframing the Judicial Diversity Debate: Personal Values and Tacit Diversity' (2015) 35(1) LS 1

—, 'The Influence of Personal Values on Legal Judgments' (2013) 40(4) J L & Soc'y 596

Cameron, CM., Park, J-K., 'How Will They Vote? Predicting the Future Behaviour of Supreme Court Nominees, 1937 – 2006' (2009) 6 JELS 485

Carrio, GR., 'Professor Dworkin's Views on Legal Positivism' (1980) 55 Ind L J 209

Casper, JD, Tyler, TR., Fisher, B., 'Procedural Justice in Felony Cases' (1988) (22) Law and Soc'y Rev 483

Chew, PK., Kelley, RE., 'Myth of the Color-Blind Judge: An Empirical Analysis of Racial Harassment Cases' (2009) (86) Wash U L Rev 1117

Cox, AB., Miles, TJ., 'Judging the Voting Rights Act' (2008) 108 Columbia L Rev 1

Crewe, B., Liebling, A., Padfield, N., Virgo, G., 'Joint Enterprise: The Implications of an Unfair and Unclear Law' (2015) Crim LR 252

Croes, MT., 'Explaining the Dealings of Dutch SMEs with Potential Legal Problems: A Plea for a Theory-Driven Approach' (2013) Int J L in Context 239

Cross, FB., 'Political Science and the New Legal Realism: A Case of Unfortunate Interdisciplinary Ignorance' (1998) 92 Nw U L Rev 251

—, FB., 'Decisionmaking in the US Circuit Courts of Appeals' (2003) 91(6) Cali L Rev 1457

—, 'The New Legal Realism and Statutory Interpretation' (2013) (1)(1) T & P of Legislation 129

—, and Tiller, EH., 'Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals' (1998) 107 Yale LJ 2155

Danziger S., Levav, J., Avnaim-Pesso, L., 'Extraneous Factors in Judicial Decisions' (2011) 108 PNAS 6889

Dennis, I., 'Fair Trials and Safe Convictions' [2003] CLP 211

—, 'The Right to Confront Witnesses: Meanings, Myths and Human Rights' (2010) 4 Crim LR 255

Devlin, P., 'The Conscience of the Jury' (1991) 107 (Jul) LQR 398

Edwards, HT., 'Public Misconceptions Concerning the 'Politics' of Judging: Dispelling Some Myths About the DC Circuit (1985) 56 U of Colorado L Rev 619

—, and Livermore, MA., 'The Pitfalls of Empirical Studies that Attempt to Understand the Factors Affecting Appellant Decisionmaking' (2009) (58)(8) Duke LJ 1895

Epstein, L., Mershon, C., 'Measuring Political Preferences' (1996) 40 (1) Am J Pol Sci 261

—, King, G., 'The Rules of Inference' (2002) 69 U Chic L Rev 1

—, WM Landes and RA Posner, 'Are Even Unanimous Decisions in the United States Supreme Court Political?' (2012) Nw U L Rev 699

—, 'Some Thoughts on the Study of Judicial Behaviour' (2016) 57 Wm & Mary L Rev 2017

Fischman, JB., Law, DS., 'What is Judicial Ideology, and How Should We Measure It?' (2009) (29) Wash U J of Law & Pol'y 133

Frank, J., 'Are Judges Human? Part 1: The effect on Legal Thinking of the Assumption that Judges Behave Like Human Beings' (1932) 80 U Pa L Rev 17

Garoupa, N., Gili, M., Gomez-Pomar, F., 'Political Influence and Career Judges: An Empirical Analysis of Administrative Review by the Spanish Supreme Court' (2012) JELS 795

George, TE., 'From Judge to Justice: Social Background Theory and the Supreme Court' (2008) 86 NCL Rev 1333

—, Yoon, AH., 'Chief Judges: The Limits of Attitudinal Theory and Possible Paradox of Managerial Judging' (2008) 61 Vand L Rev 1

Geyh, CG., 'Can the Rule of Law Survive Judicial Politics?' (2012) 97 Cornell L Rev 191

—, 'The Dimensions of Judicial Impartiality' (2013) (65) Fla L Rev 493

Green, MS., 'Legal Realism as a Theory of Law' (2005) 46(6) W & M L Rev 1915

—, 'Leiter on the Legal Realists' (2011) (30)(4) L & P 381

Greenland, S., Poole, C., 'Problems in Common Interpretations of Statistics in Scientific Articles, Expert Reports, and Testimony' (2011) (51) Jurimetrics 113

Guthrie, C., Rachlinski, JJ., Wistrich, AJ., 'Inside the Judicial Mind' (2001) (86) Cornell L Rev 777

Hagle, TM., 'Two World, Neither Perfect: A Comment on the Tension Between Legal and Empirical Legal Studies' (2015) 63 Buff L Rev 379

Hall, MA., Wright, RF., 'Systematic Content Analysis of Judicial Opinions' (2008) 96 Cali L Rev 63

Halvorsen, V., 'Is It Better That Ten Guilty Persons Go Free Than One Innocent Person Be Convicted?' (2004) 23 Crim Just Ethics 3

Hanretty, C., 'The Decisions and Ideal Points of British Law Lords' (2013) 43 (3) Brit J of Pol Sci 703

Harvey, A., Woodruff, MJ., 'Confirmation Bias in the United States Supreme Court Judicial Database' (2011) (29)(2) J of L, Econ & Org 414

Heise, M., 'The Past, Present, and Future of Empirical Legal Scholarship: Judicial Decision Making and the New Empiricism' (2002) U Ill L Rev 819

—, 'An Empirical Analysis of Empirical Legal Scholarship Production, 1990-2009' (2011) U Ill L Rev 1739

Ho, DE., 'Foreword: Conference Bias' (2013) (10) 4 JELS 603

—, Rubin, DB., 'Credible Causal Inference for Empirical Legal Studies' (2011) (7) Ann Rev of Law and Soc Sci 17

Hohfeld, WN., 'Some Fundamental Legal Conceptions As Applied in Judicial Reasoning' (1914) (23) Yale LJ 16

Holmes, OW., 'Path of the Law' (1897) 1 Boston Law School Magazine 1

Jebb, AT., Parrigon, S., Woo, SE., 'Exploratory Data Analysis as a Foundation of Inductive Research' (2017) HR Man Rev 265

Juliano, A., Schwab, SJ., 'The Sweep of Sexual Harassment Cases' (2001) (86) Cornell L Rev 548

Kahneman, D., 'A Perspective on Judgment and Choice: Mapping Bound Rationality' (2003) American Psychologist, 697-720

Kemp, V., Pleasence, P., Balmer, NJ., 'Children, Young People and Requests for Police Station Legal Advice: 25 Years On From PACE' (2011) Youth Justice 11

Kim, PT., 'Deliberation and Strategy on the United States Courts of Appeals: An Empirical Exploration of Panel Effects' (2008) 157 Pa L Rev 319

Kirby, M., 'Judicial Recusal: Differentiating Judicial Impartiality and Judicial Independence' (2015) 4 Brit J of American LS

Landes, WM., Posner, RA., 'Rational Judicial Behaviour: A Statistical Study' (2009) (1) Journal of Legal Analysis 775

Leigh, LH., 'Lurking Doubt and the Safety of Convictions' (2006) Crim LR 809

Leiter, B., 'Legal Formalism and Legal Realism: What is the Issue?' (2010) LT 111

Llewellyn, KN., 'Some Realism about Realism: Responding to Dean Roscoe Pound' (1931) (44) Harvard L Rev 1222

Loevinger, L., 'Jurimetrics: The Methodology of Legal Inquiry' (1963) 28 L & Contemp'ry Prob 5

—, 'Jurimetrics: The Next Step Forward' (1949) 33 (5) Minn L Rev 455

Lucy, W., 'The Possibility of Impartiality' (2005) 25 OJLS 3

Malleson, K., 'Appeals Against Conviction and the Principle of Finality' (1994) 21 JL & Soc'y 151

—, 'Justifying Gender Equality on the Bench: Why Difference Won't Do' (2003) (11)(1) Fem LS 1

Miles, T.J., Sunstein, C.R., 'The New Legal Realism' (2008) 75 U Chicago L Rev 831

Mitchell, G., 'Taking Behavioralism too Seriously? The Unwarranted Pessimism of the New Behavioral Analysis of Law' (2002) 43 Wm & Mary L Rev 1907

Moore, U., Callahan, C.C., 'Law and Learning Theory: A Study in Legal Control' (1943) (53)(1) Yale LJ 1

Northrop, F.S.C., 'Underhill Moore's Legal Science: Its Nature and Significance' (1950)(59) Yale LJ 196

Nuzzo, R., 'Scientific Method: Statistical Errors' (2014) (506) Nature 150

—, 'Fooling Ourselves' (2015) (526) Nature 182

Paternoster, R., Bachman, R., Brame, R., Sherman, L.W., 'Do Fair Procedures Matter? The Effect of Procedural Justice on Spouse Assault' (1997) 31 Law & Soc'y Rev 163

Peduzzi, P., Concato, J., Kemper, E., Holford, T.R., Feinstein, A.R., 'A Simulation Study of the Number of Events Per Variable in Logistic Regression Analysis' (1997) 49 (12) J Clin Epidemiology 1373

Peresie, L.J., 'Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts' (2005) 114 Yale LJ 1759

Pettys, T.E., 'Free Expression, In-Group Bias, and the Court's Conservatives: A Critique of the Epstein-Parker-Segal Study' (2015) 63 Buff L Rev 1

Pina-Sanchez, J., Linacre, R., 'Sentence Consistency in England and Wales: Evidence from the Crown Court Sentencing Survey' (2013) 53 Brit J Criminol 1118

Posner, E.A., De Figueiredo, M.F.P., 'Is the International Court of Justice Biased?' (2005) 34 J Legal Stud 599

Posner, R.A., 'The Problematics of Moral and Legal Theory' (1998) (111) (7) Harv L Rev 1637

—, 'The Role of the Judge in the Twenty-First Century' (2006) BUL L Rev 1049

—, 'Realism About Judges' (2011) Nw U L Rev 577

Pritchett, CH., 'Divided Supreme Court, 1944-1945 (1945) 44 Mich L Rev 427

Quinn, KM., 'The Academic Study of Decision Making on Multimember Courts' (2012) 100 Cal L Rev 1493

Rachlinski, JJ., Wistrich, AJ., 'Judging the Judiciary by Numbers: Empirical Research on Judges' (2017) 13 Annu Rev Law Soc Sci (forthcoming)

Rackley, E., 'When Hercules Met the Happy Prince: Re-Imagining the Judge' (2005) (12) Tex Wesleyan L Rev 213

—, 'In Conversation with Lord Justice Etherton: Revisiting the Case for a More Diverse Judiciary' (2010) PL 655

Resnik, J., 'Managerial Judges' (1983) 67 Harv L Rev 374

Revesz, RL., 'Environmental Regulation, Ideology, and the DC Circuit' (1997)(83) Virginia Law Review 1717

Roberts, S., 'The Royal Commission on Criminal Justice and Factual Innocence: Remedying Wrongful Convictions in the Court of Appeal' (2004) (1)(2) JJ 86

—, and Weathered, L., 'Assisting the Factually Innocent: The Contradictions and Compatibility of Innocence Projects and the Criminal Cases Review Commission' (2009) 29 OJLS 43

Robertson, D., 'Judicial Ideology in the House of Lords: A Jurimetric Analysis' (1982) 12 Brit J Pol Sci 1

Robinson, Z., 'Comparative Judicial Attitudinalism: A Preliminary Study of Judicial Choices in Westminster Legal Systems' (2011) U Chi Legal F 209

Schauer, F., 'Legal Realism Untamed' (2013) Tex LR 749

Schlegel, JH., 'American Legal Realism and Empirical Social Science: From the Yale Experience' (1979) (28) Buff L Rev 459

Schubert, G., 'Behavioral Jurisprudence' (1967) 2 Law & Soc'y Rev 407

Seaborne-Davies, D., 'The Court of Criminal Appeal: The first forty years' (1951) 1 (6) J Soc Pub Tchrs of Law [New Series] 425

Segal, JA., Cover, AD., 'Ideological Values and the Votes of US Supreme Court Justices' (1989) (83) Am Pol Sci Review 557

—, and Spaeth, HJ., 'Response to Critics', (1994)(4) Law and Courts 10

Shapiro, C., 'Coding Complexity: Bringing Law to the Empirical Analysis of the Supreme Court' (2009) Hastings LJ 477

—, 'Numbers, Motivated Reasoning, and Empirical Scholarship' (2015) 63 Buff L Rev 385

Sisk, GC., Heise, M., 'Judges and Ideology: Public and Academic Debates About Statistical Measures' (2005) (99) NWU L Rev 743

—, Heise, M., Morriss, AP., 'Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning' (1998) (73) NYU L Rev 1377

Skinns, L., 'I'm a detainee; Get Me Out of Here' (2009) 49 Brit J Criminol 399

Smith, JC., 'The Criminal Appeal Act 1995: Part 1: Appeals Against Conviction' (1995) Crim LR 920

Smith, R., 'Symposium: The Supreme Court and the Attitudinal Model' (1994) 4 Law and Courts Newsletter 8

Spector, PE, et al, 'Moving the Pendulum Back to the Middle: Reflections on and Introduction to Inductive Research Special Issue of Journal of Business and Psychology' (2014) J Bus Psychol 499

Spencer, JR., 'Criminal Law and Criminal Appeals: The Tail that Wags the Dog' [1982] Crim LR 260

—, 'The Sexual Offences Act 2003' (2004) Crim LR 347.

—, 'Does Our Present Criminal Appeal System Make Sense?' (2006) Crim LR 677

—, 'Quashing Convictions for Procedural Irregularities' (2007) Crim LR 835

Steyn, J (Lord)., 'Democracy, the Rule of Law, and the Role of Judges' (2006) EHRLR 243

Stribopoulos, J., Yahya, MA., 'Does a Judge's Party of Appointment or Gender Matter to Case Outcomes? An Empirical Study of the Court of Appeal for Ontario' (2007) (45) Osgoode Hall LJ 315, 347

Sunshine, J., Tyler, TR., 'The Role of Procedural Justice and Legitimacy in Shaping Support for the Police' (2003) (37) Law and Soc'y Rev 513

Sunstein, CR., Schkade, D., Ellman, LM., 'Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation' (2004) (90) Virginia L Rev 301

Thomas, C., 'How Should We Go About Jury Research in Scotland? A Response' (2016) Crim LR 915

Tukey, JW., 'We Need Both Exploratory and Confirmatory' (1980) 34 Am Stat 23

Twining, W., 'Talk About Realism' (1985) (60)(30) NYU L Rev 329

Ulmer, SS., 'Social Background as an Indicator to the Votes of Supreme Court Justices in Criminal Cases: 1947 – 1956 Terms' (1973) Am J Pol Sci 622

Vittinghoff, E., McCulloch, CE., 'Relaxing the Rule of Ten Events Per Variables in Logistic and Cox Regression' (2007) (165) Am J Epidemiology 710

Voeten, E., 'The Impartiality of International Judges: Evidence from the European Court of Human Rights' (2008) 102 (4) Am Pol Sci Rev 417

Wald, P., 'A Response to Tiller and Cross' (1999) 99 Columbia L Rev 235

Wendel, WB., 'Impartiality in Judicial Ethics: A Jurisprudential Analysis' (2012) 22 Notre Dame LJ Ethics & Pub Pol'y 305

Weinshall-Margel, K., 'Attitudinal and Neo-Institutional Models of Supreme Court Decision Making: An Empirical and Comparative Perspective from Israel' (2011) JELS 556

White, GE., 'From Realism to Critical Legal Studies: A Truncated Intellectual History' (1987) (40) SW LJ 819

Whittington, KE., 'Once More Unto the Breach: Post-Behavioralist Approaches to Judicial Politics' (2000) (25) L& Soc Inquiry 601

Williams, MS., Law, AO., 'Understanding Judicial Decision Making in Immigration Cases at the US Courts of Appeals' (2012) 33(1) Justice System Journal 97

Wilson, W., Ormerod, D., 'Simply Harsh to Fairly Simple: Joint Enterprise Reform' (2015) Crim LR 3

Winkelman, DA., et. al. 'An Empirical Method for Harmless Error' (2014) (46) Ariz State L J 1405

Winship, C., Morgan, SL., 'The Estimation of Causal Effects from Observational Data' (1999) (25) Ann Rev of Sociology 659

Wistrich, AJ., Guthrie, C., Rachlinkski, JJ., 'Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding' (2005) 4 U Penn L Rev 1251

Yung 'Flexing Judicial Muscle: An Empirical Study of Activism in the Federal Courts' (2011) (105) NWU LR 1

Cases cited

A v Secretary of State for the Home Department 2004 UKHL 56

Attorney General's Reference (No 3 of 1999) [2001] 1 AC 91

Chan Wing-Siu v R [1985] AC 168

Chevron USA Inc v Natural Resources Defense Council Inc (1984) 467 US 837

Dial and Dottin v Trinidad and Tobago [2004] UKPC 4

Findlay v United Kingdom (1997) 24 EHRR 221

GMC v Professor Sir Roy Meadow v HM Attorney General [2006] EWCA Civ 1390

Gregory v United Kingdom (1998) 25 EHRR 577

Hanif and Khan v United Kingdom (2012) 55 EHRR 16

Hui Chi-Ming v R [1992] 1 AC 34

Incal v Turkey (22678/93, 09/06/1998, Grand Chamber)

Le Compte v Belgium (1982) 4 EHRR 1

McIlkenny & Others v Chief Constable of West Midlands Police and another [1980] QB 283

Metropolitan Railway Co. v Wright (1886) 11 App Cas 152

Porter v Magill [2002] 2 AC 357

Pullar v United Kingdom (1996) 22 EHRR 391

R (on the application of E) v JFS Governing Body [2009] UKSC 15

R (on the application of Heath) v Home Office Policy and Advisory Board for Forensic Pathology [2005] EWCA Crim 1793

R (on the application of Jackson) v Attorney General 2005 UKHL 56

R v Ahmed [2010] EWCA Crim 2899

R v Ahmed [2014] EWCA Crim 619.

R v Cannings [2004] EWCA Crim 1

R v Anwar [2007] EWCA Crim 3226

R v Azam and others [2010] EWCA Crim 226

R v B [2008] EWCA Crim 559

R v Barker [2006] EWCA Crim 3249

R v Beedall [2007] EWCA Crim 23

R v Boreman and others [2006] EWCA Crim 2265

R v Bow Street Metropolitan Stipendiary Magistrate ex parte Pinochet (Pinochet No 1) [2000] 1 AC 61,

R v Bow Street Metropolitan Stipendiary Magistrate ex parte Pinochet (Pinochet No 2) [2000] 1 AC 119

R v Bow Street Metropolitan Stipendiary Magistrate ex parte Pinochet (Pinochet No 2) [2000] 1 AC 147

R v Broadhead [2006] EWCA Crim 1705

R v Brown (1910) 4 Cr App R 104

R v Castree [2008] EWCA Crim 1866

R v CCRC ex parte Pearson (2000) 1 Cr App R 141

R v Chalkley [1998] 2 Cr App R 79

R v Charova [2008] EWCA Crim 1767

R v Clark [2003] EWCA Crim 1020

R v Cooper (1969) 53 Cr App R 82

R v Cordingley [2007] EWCA Crim 2174

R v Dada [2008] EWCA Crim 2121

R v Dass [2009] EWCA Crim 1208

R v Davis, Rowe, and Johnson [2001] 1 Cr App R 115

R v Erskine; R v Williams [2009] EWCA Crim 1425

R v Evans (1950) 34 Cr App R 72

R v F [2009] EWCA Crim 805

R v Fletcher [2009] EWCA Crim 1187

R v Forbes [2001] 1 AC 473

R v G [2009] 1 AC 92

R v Garland [2016] EWCA Crim 1743

R v Gibson [2006] EWCA Crim 542

R v Gough [1993] AC 646.

R v Haigh [2010] EWCA Crim 90

R v Hanrahan [2015] EWCA Crim 1653

R v Hanratty [2002] Cr App R 30

R v Hopkins-Husson (1950) 34 Cr App R 47

R v Horseferry Road Magistrates' Court ex parte Bennett [1994] 1 AC 42

R v Inglis [2010] EWCA Crim 2269

R v J [2010] EWCA Crim 1768

R v Jheeta [2007] EWCA Crim 1699

R v Jogee; Ruddock v R [2016] UKSC 8, [2016] UKPC 7

R v Johnson and others [2016] EWCA Crim 1613

R v Khan and Others [2008] EWCA Crim 531

R v Khokhar [2007] EWCA Crim 1756

R v L [2007] EWCA Crim 1750

R v Latus [2006] EWCA Crim 3187

R v Lawless [2009] EWCA Crim 1308

R v Martin T [2008] EWCA Crim 3229

R v Mason (1924) 17 Cr App R 160

R v McDonagh [2012] EWCA Crim 2013

R v McGrath [1949] 2 All ER 495

R v McIlkenny & Others [1991] 93 Cr App R 287

R v Michael A [2006] EWCA Crim 905

R v Mullen [1999] 2 Cr App R 143

R v O'Leary (aka Dwyer) [2006] EWCA Crim 3222

R v Paris, Adbullahi, and Miller (1993) 97 Cr App R 99

R v Parks (1962) 46 Cr App R 29

R v Pendleton [2001] UKHL 66

R v PF [2009] EWCA Crim 1086

R v Pope [2012] EWCA Crim 2241

R v Powell; R v English [1999] 1 AC 1

R v Puaca [2005] EWCA Crim 3001

R v Ramirez [2009] EWCA Crim 1721

R v Reed and Reed; R v Garmson [2009] EWCA Crim 2698

R v Robert (A.K.A. Sean) Hogdson [2009] EWCA Crim 490

R v Rowland (1948) 32 Cr App R 29

R v Silcott 1991 Times December 9

R v Simmons [2009] EWCA Crim 741

R v Simpson (1909) 2 Cr App R 128

R v Smith [1999] 2 Cr App R 238

R v Stafford; R v Luvaglio (1969) 53 Cr App R 1

R v Stanley [2008] EWCA Crim 603

R v Sussex Justices ex p. McCarthy [1924] 1 KB 256

R v Tellett (1921) 15 Cr App R 159

R v Togher; R v Doran; R v Parsons [2001] 1 Cr App R 33

R v Weir [2001] 1 WLR 421

R v Weir Times 16 June 2000

R v Williams [2007] EWCA Crim 2264

R v Williamson (1909) 1 Cr App R 3

R v Wood [2008] EWCA Crim 587

Radmacher v Granatino [2010] UKSC 42

Scruttons Ltd v Midland Silicones Ltd [1962] AC 446

Stafford v DPP [1974] AC 878

Stirland v DPP [1944] AC 315

Woolmington v DPP [1935] AC 462

Legislation

Constitutional Reform Act 2005

Criminal Appeal Act 1907, 1964, 1966, 1968, 1995

Criminal Procedure Rules 2016

European Convention on Human Rights

Homicide Act 1957

Hunting Act 2005

Magistrates Court Act 1980

Parliament Act 1911, 1949

Police and Criminal Evidence Act 1984

Reports, official documents, theses, etc.

An Overview of Sexual Offending in England and Wales (Home Office, Ministry of Justice 2013)

Berry, DA., 'P-Values Are Not What They're Cracked Up To Be' a contribution to the ASA statement roundtable, submission No 5 (ASA 2016)

Departmental Committee on New Trials in Criminal Cases, *Report*, (Cmd. 9150, 1954)

Genn, H., B Lever, L Gray, *Tribunals for Diverse Users: DCA Research Series No1/2006* (DCA 2006)

Greenland, S., et al 'Statistical Tests, P-Values, Confidence Intervals, and Power: A Guide to Misinterpretations' <http://amstat.tandfonline.com/doi/abs/10.1080/00031305.2016.1154108>, supplemental submission No 21

Heaton, SJ., *A Critical Evaluation of Using Innocence as a Criterion in the Post-Conviction Process* (DPhil Thesis, University of East Anglia, 2013)

HM Courts Service, *A Guide To Commencing Proceedings in the Court of Appeal Criminal Division* (HMCS 2008)

Homicides, Firearm Offences, and Intimate Violence 2009-10 (Home Office 2011)

House of Commons Justice Select Committee, Criminal Cases Review Commission, 12th Report of 2014–15, (March 2015, HC 850)

Howard League, *Judging Maturity: Exploring the Role of Maturity in the Sentencing of Young Adults*, (Howard League 2017)

In the Court of Criminal Appeal: Annual Report 2014/15 (Court of Appeal 2015)

In the Court of Appeal Criminal Division Annual Report 2015-16 (Court of Appeal 2017)

Independent Police Complaints Commission, *Destruction of Specific Documents Leading to the Collapse of the R v Mouncher & others trial*, *Report*, (IPCC 2011)

Interdepartmental Committee on the Court of Criminal Appeal, *Report* (Cmnd 2755, 1965);

Jackson, J., et al, 'European Social Survey: Trust in Justice Topline Results from Round 5 of the European Social Survey' (European Social Survey 2011)

Jansson, K., 'Public Confidence in the Criminal Justice System: Findings from the Crime Survey for England and Wales 2013/14' (Ministry of Justice 2015)

JUSTICE Committee, *Criminal Appeals* (Stevens and Sons 1964)

Lew, MJ., 'Three Inferential Questions, Two Types of P-Value', submission 14 to ASA Statement on P-Values (ASA 2016)

Morris, S., 'Cardiff Three Are Completely Innocent, Says Convicted Killer of Lynette White' (The Guardian, 13 July 2011)

Roberts, S., *The Decision Making Process of Appeals Against Conviction in the England and Wales Court of Appeal (Criminal Division)* (DPhil thesis, London School of Economics and Political Science, 2009)

Royal Commission on Criminal Justice, *Report* (cm 2263, 1993)

Thomas, C., *Judicial Diversity in the United Kingdom and Other Jurisdictions* (Commission for Judicial Appointments 2005)

—, *Are Juries Fair?* (MOJ 2010)

—, H Genn 'Understanding Tribunal Decision-Making: A Foundational Empirical Study (Preliminary Report)' (Nuffield Foundation 2013),

—, *Judicial Diversity in the United Kingdom and Other Jurisdictions: A Review of Research, Policies and Practices* (Commission for Judicial Appointments 2005),

Wasserstein, RL., Lazar, NZ., 'The ASA's Statement on P-values: Context, Process and Purpose' (American Statistical Association, 2016)

Websites cited

<http://www.artsandsciences.sc.edu/poli/juri/appct.htm> <accessed 26 July 16>

<http://www.artsandsciences.sc.edu/poli/juri/appct.htm> <accessed 8 August 2016>

<http://www.artsandsciences.sc.edu/poli/juri/databases.htm> <accessed 26 July 16>

<http://www.artsandsciences.sc.edu/poli/juri/highcts.htm> <accessed 26 July 16>

<http://www.ccrcc.gov.uk/case-statistics/> <6 June 2016>

<http://www.gov.uk/government/statistics/court-statistics-quarterly-january-to-march-2014> <accessed 29 July 2016>.

<https://www.gov.uk/government/statistics/court-statistics-quarterly-earlier-editions-in-the-series> (accessed 19/06/2017)

<https://www.gov.uk/government/statistics/criminal-court-statistics-quarterly-october-to-december-2016> (accessed 19/06/2017)

<http://www.law.cf.ac.uk/newsandevents/transcripts/271005.pdf> <accessed 8 November 2015>.

<http://www.news.bbc.co.uk/1/hi/wales/7792501.stm>. <18 August 2016>

<http://www.supremecourt.gov/faq.aspx#faqgi1> <accessed 20 August 2015>

<http://www.supremecourtdatabase.org/>

<https://www.judiciary.gov.uk/about-the-judiciary/who-are-the-judiciary/diversity/judicial-diversity-statistics-2015/> <accessed 6 September 2016

<https://www.judiciary.gov.uk/you-and-the-judiciary/going-to-court/court-of-appeal-home/coa-biogs/> <accessed 20 November 2015>

<https://www.sentencingcouncil.org.uk/analysis-and-research/crown-court-sentencing-survey/> <10 September 2016>

<https://www.supremecourt.uk/docs/speech-151106.pdf> <accessed 6 September 2016>