The colour of Justice: Does race influence the decisions of criminal justice workers in England and Wales?

DONNA SPENCER, BSc, PGCE, MSc Applied Forensic Psychology

Thesis submitted to the University of Nottingham for the degree of Doctor of Forensic Psychology

2022

Submitted for Examination July 2022

Abstract

Background: Black people in the UK are disproportionately arrested, prosecuted, and remanded in custody. This has led some people to question whether the criminal justice system (CJS) is racially biased. Some juror research has found that the race of the defendant and victim may influence jurors' decisions on verdict, sentencing and blame. However, few studies have been done in England and Wales and using criminal justice workers (CJWs). Studying CJWs is important as they are involved in all stages of the criminal justice process.

Aim: This doctoral research investigates the possibility of racial bias in guilt perception, blame attribution, and sentencing decisions by CJWs in England and Wales.

Method: The thesis consists of three main studies. A systematic review was conducted on papers published between 2015 to 2021 that investigated the influence of defendants' and victims' race on mock jurors' decisions regarding guilt, blame, and sentencing. Two cross sectional studies of CJWs from England and Wales were also conducted using written case vignettes in the context of a mock juror paradigm. Study 1 examined the effect of defendant and victim race and their interaction with the criminal justice profession on perceived guilt, attributed blame, and suggested punishment by 399 CJWs. Using the same methodology, independent and dependent variables, study 2 focused on White CJWs only (N = 326) to explore in-group favouritism. In both studies, the Balanced Inventory of Desirable Responding (BIDR-16) was employed to assess and control for socially desirable responding amongst participants. Thus, the thesis also included a critique of the BIDR-16.

Results: The systematic review of 28 studies of low-moderate quality, including a total of 9,963 mock jurors, did not find clear evidence of racial bias against Black or White defendants and victims in verdicts, sentencing decisions, or attributed blame by mock jurors. Instead, the decisions were influenced by the type of crime and other extra-legal factors, for example, mental health status, intellectual disability, physical attractiveness, race saliency of defendants and/or victims. Study 1 also

did not find any significant effect of defendant/victim's race on the perceived guilt, attributable blame, or suggested sentence. In comparison, CJWs' profession did have a significant effect on perceived guilt (p = .005), with probation workers ascribing more guilt to the defendants compared with youth justice/social workers. Similarly, CJWs with different professions suggested significantly different punishments when the defendant was described as Black (p = .002) or Asian (p = .005). Youth justice/social workers gave the most lenient sentences in scenarios involving a Black and Asian defendant or victim. Similarly, study 2 with White CJWs only found no significant effect of victim and defendant's race on decisions of guilt, blame, and punishment. Perceived guilt (p = .005) and suggested sentence (p < .001) were significantly affected by CJWs' profession. Probation workers ascribed a higher degree of guilt to the defendant than youth justice/social workers (p = .006). Also, compared with youth justice/social workers, police (OR = 5.03, 95%CI [2.12, 11.94]), probation (OR = 8.13, 95%CI [1.50, 12.75]), and prison workers (OR = 4.36, 95%CI [3.06, 21.58]) were more likely to choose prison sentence over no punishment. The critique of the BIDR-16 scale showed internal and temporal consistency. However, the reliability of its SDE subscales was questionable (Cronbach's α, .63 – .66), though its IM subscale's reliability was fairly acceptable (Cronbach's α, .71 – .74).

Conclusion: This thesis did not find clear evidence of racial bias in terms guilt perception, blame attribution, and sentencing decisions among mock jurors, including CJWs in England and Wales. In contrast, mock jurors' decisions appear to be driven by extra-legal factors such as CJWs' profession, and non-racial characteristics of the defendants. Variations in mock jurors' decisions on the crimes committed by Black Asian and Minority Ethnic (BAME) versus White defendants could reflect real differences in the types of crime committed by BAME versus White defendants; perhaps certain BAME groups have intersectionality differences that impact on crime; or maybe racial bias is occurring elsewhere in the CJS. Future research should consider how decisions made at other stages of the CJS might account for racial disproportionality, explore ingroup bias amongst BAME CJWs, and use case examples more applicable to the job role of the CJW.

Table of Contents

Abstract 2

List of Tables 5

List of Figures 7

Acknowledgements 8

Chapter 1: General introduction 9

Chapter 2: Do mock-jurors treat Black defendants and victims differently from White defendants and victims on decisions of guilt, blame and punishment? A systematic review. 28

Chapter 3: Measuring Socially Desirable Responding: A Critique of the BIDR-16 75

Chapter 4: The influence of defendant and victim race on Criminal Justice Workers' decisions on guilt, punishment, and blame. 92

Chapter 5: The influence of outgroup bias in the attribution of guilt, blame, and sentencing decisions made by White criminal justice workers in England and Wales. 127

Chapter 6: Discussion, conclusions, and future research 152

References 165

Appendices 188

List of Tables

Number	Title	P
Table 1	An example of the search strategy used to search the PsychoInfo database	32
Table 2	Study characteristics	35
Table 3	SR study results - summarises the design, outcome measures and findings from the studies in this review.	48
Table 4	Goodness-of-fit Indicators of models for a 36-item, 20-item, and 16-item two-factor BIDR model	82
Table 5	Goodness-of-fit indicators for the BIDR-16 by nation and gender	83
Table 6	Correlations between the BIDR, BIDR-16, and existing scales in study 1.	84
Table 7	Correlations between BIDR-16, SDE and IM, and existing scales in study 4.	86
Table 8	Demographic characteristics	97
Table 9	Descriptive characteristics of the entire study population	104
Table 10	Correction of socially desirable responding (as measured by the BIDR-16) with guilt and blame	105
Table 11	Perceived guilt and attributed blame by defendant and victim's race	106
Table 12	Suggested sentence by defendant and victim's race	107
Table 13	Perceived guilt and attributed blame by CJWs' professions	108

Table 14	Suggested sentencing across criminal justice professions	111
Table 15	Suggested sentence across criminal justice professions by defendant's race	112
Table 16	Suggested sentence across criminal justice professions by victim's race	114
Table 17	Descriptive characteristics of White CJWs	134
Table 18	Correlation of socially desirable responding with perceived guilt and attributed blame among White CJWs	134
Table 19	Perceived guilt and attributed blame among White CJWs by defendant and victim's race	135
Table 20	Sentencing decisions by White CJWs across defendant and victim race.	137
Table 21	Perceived guilt and attributed blame by White CJWs' professions	138

List of Figures

Number	Title	p
Figure 1	Steps of the PRISMA process and the number of citations excluded at each stage	44
Figure 2	Perceived mean guilt by CJWs' profession	109
Figure 3	Attributed blame across CJWs' professions	110

Acknowledgements

I would like to thank my family (particularly my parents, Warren and Clare Spencer; and sister Michelle), my late Godfather (Captain Samuel Taylor) friends (Dr. Sacha Brostoff, Etonam), those from my cohort (Yvette, Sharron), supervisors (Dr. Vince Egan and Dr. Nigel Hunt) for helping me through this journey. Thank you also for the 399 criminal justice workers who took the time to participant in this research. I am particularly thankful to those who took the time to listen to my woes, read my work, keep me positive, and give me hope and encouragement in the times when I felt ready to quit. God bless you all. I have learnt a lot beyond compiling this thesis. This test has become my testimony. My hope is that my journey will inspire others, and my research will provide a better understanding of the impact of race on criminal justice workers decisions within the England and Wales Criminal Justice Service.

"*Is the Lord's arm too short? Now you will see whether or not what I say will come true for you*."

(Numbers 11:23. *The Bible, New King James Version*).

Chapter 1: General introduction

Background

In England and Wales, the Criminal Justice System (CJS) is a network of agencies consisting of the police, law enforcement officers, the courts, the Home Office, His Majesty's Crown Prosecution Service Inspectorate (HMCPSI), and His Majesty's Prison Service (HMPPS). The role of the CJS is to control crime and deliver penalties to those who commit crimes.

Over the past forty years, several high-profile events including the Bristol, Brixton and Toxteth riots in 1981, the police handling of the Stephen Lawrence murder in 1990, and the London riots in 2011, have raised questions in some quarters as to whether Black people particularly are fairly treated in the CJS. After the murder of Stephen Lawrence, institutional racism in the Metropolitan Police Service was subsequently highlighted in the Macpherson Inquiry report (Macpherson, 1999), which recommended a comprehensive review of policies and practices in all institutions to ensure that no communities would be discriminated against.

Almost twenty years after the publication of the Macpherson Inquiry report and following government concerns about the number of Black, Asian, and minority ethnic (BAME) people in the CJS, the UK government commissioned David Lammy to conduct an independent review on the treatment and outcomes for BAME people in the CJS. The Lammy Review (Lammy Review, 2017), alongside some other studies, reported that compared to White people and their representation in the general population, Black people are disproportionally arrested (Lammy, 2017), prosecuted (Lammy, 2017), remanded in custody and sentenced to prison (Ministry of Justice, 2016, 2017; Office of National Statistics, 2012; Uhrig, 2016), and disproportionately receive adjudications (i.e., punishments by prison disciplinary hearings) whilst in prison (Uhrig, 2016).

Overall figures in England and Wales indicate that whilst BAME people represent 14% of the general population, they constitute 25% of the adult prison population and 40% of young people in custody (Lammy 2017). Such findings are consistent with the international findings that certain BAME groups are often overrepresented in the prison systems in Canada, USA, Australia, and New Zealand

(Australian Bureau of Statistics, 2017; New Zealand Department of Corrections, 2016; The Whitehouse (TWH), 2015; Zinger, 2017). Such racial disproportionalities in the CJS might lead some to think that BAME people generally, and Black people specifically, are unfairly treated. It is perhaps unsurprising that recent research has indicated that a majority of BAME people in England and Wales believed that the CJS discriminates against them (Centre for Justice Innovation, 2017).

However, it is worth noting that racial disproportionality in the CJS is a complex area. When the BAME category is broken down, there are differences in the representation of BAME subgroups in the CJS across England and Wales. Specifically, there is an overrepresentation of Black people within young offenders' institutions (YOIs) and prisons compared with their representation in the general population (i.e., YOIs 20% and Prisons 12% versus 3% representation in the general population), and an overrepresentation of people of mixed ethnic heritage (Lammy, 2017). Contrastingly, Asian people are underrepresented in the prison population compared to their representation in the general population (Lammy, 2017). Furthermore, it is important to note that disproportionality also varies by offence type (see Thomas, 2010).

The mixed racial disproportionality figures within the BAME category suggest it is important to explore the factors that could be influencing such discrepancies. Whilst this discrepancy might indicate a systemic racial bias against certain racial groups in the CJS, it is possible that these racial disparities reflect true differences in the amount, type, and severity of crime committed by BAME people compared with White people, rather than a racially biased CJS (Bridges, et al., 1987; Lammy, 2017). Furthermore, being of a specific BAME background might also correlate with other risk factors for offending, including low socioeconomic status (Farrington 1992a, 1992b), low parental supervision (Smith & Stern, 1997), exposure to parental conflict and interpersonal violence (Buehler et al., 1997; Fergusson & Horwood, 1998), and living in deprived areas and in poor housing (Thornberry et al., 1995). Consequently, the reasons for racial disproportionality in the CJS seem both complex and

complicated to disentangle, with non-racial individual and environmental factors potentially contributing to such disproportionality.

Race and criminal justice research: Juror and Jury research.

An area of criminal justice research where the treatment of BAME people has been extensively studied, but mostly outside of the UK, is jury and juror decision making. Note the limited research conducted in England and Wales means that studies from elsewhere will be considered, in as far as they provide a general overview of the topic. It is not intended to use findings from elsewhere to generalise to the decisions made by juries and jurors England and Wales. Rather, it is done to provoke thought that if there are racial biases, these also might occur in England and Wales and with a specific criminal justice profession (CJP) and amongst criminal justice workers (CJWs). A criminal justice professional is someone who works within the CJS, whose role it is to help deliver justice by ensuring that people who break the law are held accountable. In the context of this thesis, a criminal justice professional includes Police Officers, Probation Officers, Prison Officers, Youth Justice Social Workers, Forensic Psychologists, Lawyers, Social Workers working within the context of the HMPPS who are involved in making decisions about suspects/ offenders/ prisoners (e.g., arrest, release from prison). Studies from juror and jury research are discussed here primarily as the simulated method was used as part of this thesis to explore the impact of defendant and victim race on CJWs' decisions on guilt, punishment, and blame.

Note, studies in the area have tended to compare decisions on Black and White defendants and victims, which might largely be because when the BAME category has been broken down elsewhere (e.g., the United States), it is Black (and Hispanic) people that have tended to be disproportionately represented in various CJSs compared with other racial groups. Very little research has been conducted examining decision making around underrepresented racial groups, particularly with Asian defendants and victims. The lack of studies on the influence Asian defendant and victim race on juror and jury decisions has meant there is little discussion of this group here. However, the lack of research

on decisions regarding Asian defendants and victims highlights the need for research to be conducted with this racial group. Indeed, this gap in research has contributed to the decision to include in this thesis, an exploration of whether being an Asian defendant or victim influences criminal justice workers decisions on guilt, sentencing and blame in the current thesis.

The subsequent discussion in this chapter will initially describe the archival and simulated methods. Then the outcome measures used to assess racial bias in jury and juror decision making will be examined, alongside the issue of possible risk of bias associated with administering self-report inventories in research, and ways to combat this possibility. Then a discussion on the evidence from simulated and archival research will be outlined, followed by the limitations of juror and jury research. Finally, an outline of the rationale for the current thesis will be provided.

Methodology used to study jury and juror decisions.

A jury comprises a small group of demographically representative lay people whose role it is to individually listen to the evidence presented in a trial, and following legal insights of a judge, collectively deliberate over the testimonies and evidence presented and decide on the innocence/guilt of the defendant(s).

The direct observation of jury and juror decision making is often legally prevented, so innovative methods have been developed to study their decisions and decision-making processes, including examining archival data, and devising simulated trials.

The archival method. Archival research involves the researcher examining court manuscripts, and the verdicts and sentencing decisions of past cases. Consequently, the researcher can compare whether there are discrepancies in decisions made across different racial groups. Note, archival research focuses on decisions made by juries (i.e., group decisions) rather than individual jurors.

Archival research has the advantage of high ecological validity due to being based on verdict and sentencing decisions from actual cases, and can cover an extensive period, allowing for longitudinal

analysis. However, this retrospective method has some limitations. An important factor to consider when interpreting archival data is the law at the time when decisions were made. For example, archival research conducted in the US by Bullock (1961) reported that Black defendants were more likely to be sentenced than White defendants. However, at the time of that study, Jim Crow laws legalised racial segregation, whereby any breach in the law would result in fines or imprisonment for Black people. Therefore, evidence of bias against Black defendants/ offenders may not reflect racial biases held by jurors/juries, rather it could be indicative of the laws effective at the time (Kennedy, 2012). In contrast, there have never been any laws permitting racial discrimination in the UK and, therefore, cross-cultural comparisons should be made with caution. Indeed, different countries have different laws, legal systems, and records on race relations. Nevertheless, speculatively speaking, Britain's involvement in the transatlantic slave trade and its colonial past might have influenced views about certain BAME groups relationship with crime, criminality, and ultimately their treatment within the CJS in post-colonial Britain. Other disadvantages of the archival studies include the possibility of missing information (Kapardis, 1997) and issues with data access (Bray & Kerr, 1982), time-consuming process of data extraction, limited reproducibility of the data, and the lack of experimental control to adjust for other contributing variables.

The simulated method. Due to archival research limitations, simulated methods have sometimes been used. The simulated method, also known as the mock jury/juror paradigm, involves presenting a mock trial (via video, audio, written vignette) to a mock juror/jury (i.e., participants), varying the characteristics of the defendant(s), the victim(s), mock juror(s), or evidence, whilst keeping the other facts of the case constant. The researcher then observes the effects of the manipulation (e.g., race) on mock jurors'/juries' decisions on outcomes such as verdict, sentencing and blame (see Landy & Aronson, 1969). Simulated studies have examined how the decisions of juries (i.e., group decision-making) and jurors (i.e., individual decision-making) are affected by racial combinations of defendant/offender, victim, and the decision-makers themselves (Landy & Aronson, 1969). Simulated research can focus on decisions made by juries and individual jurors. The importance of juror research

is rooted in the observation that jurors often decide the defendant's innocence or guilt at the beginning of a trial, and that these impressions may impact on eventual jury verdicts (Jones & Kaplan, 2003).

An advantage of the simulated method is the ability to introduce experimental control, increasing the internal validity. Through using experimental control, it is possible to establish the possibility of causal relationships which can advance our understanding of how extra-legal factors such as race can influence decisions. Moreover, hypotheses can be easily tested, and studies can be easily replicated using this method. The method can help identify how both juries and jurors weigh evidence, process information, and arrive at decisions. However, limitations of the simulated method include its use of transcripts, and audio recording and video clips, which is very different from a real trial setting, thereby reducing its ecological validity (Bornstein et al., 2017).

Outcome measures

Both archival and simulated studies often use verdict and sentencing decisions as outcome measures, either continuously or dichotomously (Devine & Caughlin, 2014; Mitchell et al., 2005). In the context of juror decision making, a dichotomous measure is a measure that takes on just two values, such as guilty/not guilty; whilst a continuous measure requires the respondent to rate on a numbered scale the extent that they think the defendant is guilty of an offence, for example, from not at guilty (1) to guilty (7). Blame attribution has been measured less often, either dichotomously or continuously, and exclusively in simulated research. Simulated research has the advantage of probing into the underlying reasons for verdict and sentencing decisions by jurors, such as exploring the role blame might play in decisions.

Note, many simulated studies have employed self-report measures to assess participants' judgements of guilt/verdict, sentencing and blame. Consequently, there is some concern about the risk of bias associated with self-reported outcomes (Leite & Beretvas, 2005). An important source of information bias in such research designs is the concept of socially desirable responding (SDR).

Addressing SDR in self-report inventories

SDR is a form of response bias where the respondent is inclined to provide positive self-descriptions (Johnson & Fendrich 2005). Several assessment tools have developed ways to address SDR, including standalone SDR psychometric tests (e.g., Balanced Inventory of Desirable Responding (BIDR) by Hart et al., 2015), SDR scales incorporated in other psychometric constructs (e.g., personality measures), and rational techniques (e.g., demand reduction methods) when administering self-report questionnaires (Paulhus & Vazire, 2007). A positive correlation between the SDR measures and the outcomes of interest would then be an indication of SDR and responding bias in the study (Nederhof, 1985).

Indeed, some studies have found that over the years racist attitudes and beliefs have become more covert, perhaps triggered by various race equality legislations, positive discrimination practices, and a possible desire for people to be or appear egalitarian (Dividio & Gaertner, 2000). Therefore, people who possibly hold racially biased beliefs and attitudes might suppress these views when completing self-report questionnaires (Axt et al., 2016). It therefore seems necessary that a SDR measure is used when administering self-report questionnaires, particularly those that cover sensitive topics such as racially biased attitudes and beliefs.

Findings from juror and jury research

Regardless of whether jury or juror decisions are examined through archival or simulated methods, the expectancy is that their decisions should be based on evidential facts of the case. However, research, particularly simulated research, has tended to find that their decisions are influenced by extra-legal factors including the personal characteristics of the juror, defendant(s), victim(s), and witness(es) (Devine & Caughlin, 2015).

Effect of defendant race on decisions

Guilt/verdict and sentencing. There is limited archival and simulated research conducted in England and Wales into jury decision making. A study by Thomas (2010) used archival data to examine whether jury verdicts in all Crown Courts in England and Wales between 2006-2008 correlated with defendant ethnicity, type and severity of offence and the court of hearing. Consistent with disproportionality figures, Thomas found a slightly higher conviction rate for Black defendants as compared to White and Asian defendants (67% vs 63%). In comparison Thomas' (2010) simulated research found no significant difference in the verdict decisions for White, Black, and Asian defendants by mock jurors. The simulated finding suggests that at least in terms of jury decisions in the England and Wales CJS, decisions appeared to be operating unbiasedly. However, these are only two studies, making it is difficult to make generalisations from them. Furthermore, the studies were conducted some time ago and decision-making behaviour might have changed over time. Reasons for changes in behaviours might include people becoming less racially biased or less inclined to overtly express biased beliefs (Axt et al., 2016). Consequently, more simulated and archival studies need to be conducted in England and Wales to get a better understanding of the influence of race of defendant on juror and jury decisions.

Contrastingly, the international simulated studies that have examined decisions for Black versus White defendants are mixed. Some studies have found Black defendants were more likely to be judged as guilty and recommended longer sentences than White defendants (e.g., Maeder et al. 2015, study 2), perhaps linking with US and Canadian archival research findings that Black people are often associated with criminality and dangerousness (Eberhardt et al., 2004; Singh & Sprott, 2017). However, some US and Canadian research found that race did not significantly impact on juror verdicts (e.g., Ewanation & Maeder 2018; McManus et al., 2018); whilst others found that mock jurors were significantly more likely to judge a White than Black defendant as guilty (Pica et al., 2017). Differently also, some research has indicated that compared to other minority groups (i.e., Black and Latino) and White people, Asian

American defendants tended to be treated more leniently in terms convictions and sentencing decisions, including being less likely to be imprisoned and more likely to receive shorter sentences (Franklin & Fearn, 2015; Johnson & Betsinger, 2009; Kutateladze et al., 2014).

Several possible explanations can be given for the conflicting results including: differences could reflect differences in the type of crime used across studies (Devine & Caughlin, 2015); differences in the types of measurement scales used across studies, as some studies used dichotomous whilst others continuous measurements (Devine & Caughlin, 2015); possible cultural differences between countries in terms of the influence of race of defendant on juror/jury decisions; and/or changes in attitudes and behaviours by jurors/juries over time. Importantly also, it is worth noting that defendant's race has been found to interact with other extra-legal factors, including socioeconomic status (SES) and physical attractiveness, to affect mock jurors' verdict and sentencing decisions (Armenta 2017). Perhaps then, these interactions are like the concept of intersectionality (Crenshaw, 1989), whereby it is the accumulation of someone's multiple identities (e.g., race and SES) that intersect to either advantage or disadvantage them in the CJS. Perhaps then, certain BAME racial groups are more likely to have additional identifies that disadvantage them in the CJS (e.g., being of low socioeconomic status), which collectively account for their disproportionality. Indeed, there is research indicating that mock jurors gave harsher punishments (i.e., guilty verdicts and longer sentences) to Black defendants of low SES than Black defendants of high SES and White defendants of any SES (Espinoza et al., 2015).

In terms of punishment/sentencing decisions, there is a lack of research in England and Wales. However, some US archival research has shown that juridical outcomes for Black defendants are worse than for White defendants, even after taking crime severity and criminal history into account (King & Light, 2019). Notably also, US research found that Black murderers who were facially more stereotypically Black in appearance (e.g., dark skin, broad nose, thick lips), were most likely to receive the death sentence (Eberhardt, et al., 2006); and that Judges tended to give longer prison sentences to Black offenders whose features were stereotypically Black, than to Black offenders whose features

were not stereotypically Black (Blair et al., 2004). If possible, it might be useful to conduct similar research in England and Wales, to see if being more stereotypically Black in appearance affects verdict and sentencing decisions.

Effect of victim race and juror decisions

Guilt/verdict and sentencing. There is limited archival research conducted in England and Wales in terms of studying the effect of victim race on mock juror's guilt and sentencing decisions. Thomas (2010) found that victim race did not significantly influence jury verdicts in Crown Courts in England and Wales between 2006-2008. Similarly, Thomas' (2010) simulated study found that victim race did not bias the decisions of a sample of mock jurors/juries from England and Wales.

Comparatively, some US archival research found that juridical outcomes for Black victims were worse in terms of verdict and sentencing decisions (i.e., type of sentence and length of sentence) than for White victims, even after taking crime severity and criminal history into account (King & Light, 2019; Baldus et al., 1998). Several studies have found jurors gave more severe sentences towards defendants whose victims are White (Baldus et al., 1998; Curry 2010; Eberhardt et al., 2006; Holcomb et al., 2004). The contrasting findings of these and Thomas' studies suggests that whereas victim race seems not to influence the decisions of juries in England and Wales, it does in the US. The finding suggests that juries in England and Wales are operating unbiasedly. However, there is a lack of archival research in England and Wales, and more archival studies need to be conducted before any definite conclusions can be made.

In terms of simulated research, US studies have found mixed results. There is research that found no significant difference in verdict and sentencing decisions by victim race (Call, 2019); research that reported mock jurors were significantly more likely to recommend the death penalty when the victim was Black than White (Bell, 2019); and meta-analytic research that indicated harsher punishments were given to the defendant when the victim was White (Mazella & Feingold, 1994).

Again, the variations in findings might reflect differences in the type of crime used across studies (Devine & Caughlin, 2015); cultural differences in racial attitudes between mock jurors in the US versus England and Wales; and/or changing attitudes over time. Regarding the latter point, the recent influx of studies within the area warrants another systematic review, as attitudes and behaviours might have changed with time.

The interactions between defendant and victim race.

Guilt/verdict and sentencing. The simulated study by Thomas (2010) found that White defendants accused of assaulting a BAME victim were significantly more likely to be found guilty than other defendant-victim racial combinations. There is a lack of other research in England and Wales in this area. The limited studies in this area make it difficult to draw conclusions.

Regarding international studies, there are some simulated studies indicating that guilt and punishment decisions are harshest when the offence is interracial (e.g., Hymes et al., 1993); and contrary to the finding of Thomas (2010), most specifically when the defendant is Black and victim is White (Lynch & Haney, 2000; Saucier et al., 2010; Stevenson et al., 2009). However, other simulated studies have found same-race defendant and victim pairings resulted in more guilty decisions by mock jurors (e.g., Bottoms et al., 2004; Lee et al., 2019). Again, differences in findings might reflect changes in attitudes and behaviour over time, and the type of crime. The contrasting findings perhaps indicate that more research needs to be conducted in the area.

The interactions between defendant, victim, and juror race.

Interactions between defendant, victim and juror race have generated discussion around the possible influence of ingroup and outgroup bias. An ingroup is a social group that a given individual identifies with, whilst an outgroup is a social group that a given individual does not identify with (Tajfel & Turner, 1986). A social group refers to two or more people having a shared identity based on one or more characteristics, such as race/ethnicity, gender, sexuality, political affiliation. Ingroup bias (sometimes referred to as ingroup favouritism) refers to the tendency for people to favour people who are like

them (e.g., same race) in domains including attitudes, behaviours, and perceptions (Sherif & Sherif, 1953; cf. Tajfel, 1982). Outgroup bias (or outgroup favouritism) occurs when more favourable judgements are given to members of the outgroup.

Guilt/Verdict and sentencing. Again, little archival and simulated research has been conducted in England and Wales that has explored how the race of the juror, defendant and victim might interact to influence the decisions of jurors. The study from Thomas (2010) found there was no significant difference in the treatment of Black versus White defendants in verdict decisions amongst White mock jurors. However, there was evidence of outgroup favouritism from White mock jurors who lived in more racially diverse areas. Note, there is a lack of research in England and Wales examining whether Black mock jurors show ingroup bias.

Regarding international studies, the findings are mixed. Some early simulated studies conducted in the US, Canada, and Australia, found that both White and Black jurors tended to favour the defendant or victim of their own same racial group (FosterLee et al., 2006; Pfeifer & Ogloff, 2003; Toobin, 1996). Contrastingly some other studies have found White mock jurors show outgroup bias towards Black compared to White defendants (FosterLee et al., 2006; Marcus-Newhall, et al., 2002); and other studies have found that White juror bias against Black defendants predominately occurs when the evidence against Black defendants is strong, but not when the evidence is balanced or weak (Ugwuebu, 1979; Whitestone et al., 1995).

Due to the various mixed findings, several meta-analyses have been conducted to help clarify the role that race plays in juror decision making. The meta-analyses of Mitchell and colleagues (2005) and Devine and Caughlin (2015) found that mock jurors were more likely to show leniency in terms of guilt and sentencing decisions, toward defendants of their racial ingroup. Interestingly also, Devine and Caughlin (2015) found that ingroup bias was more prominent amongst Black than White mock jurors. However, it should be noted that most of the included studies were from North America, and none included studies from England and Wales, thereby limiting the generalisability of the findings.

Again, the inconsistent findings in the area might be due to cultural differences, changes in attitudes over time, type of crime, the makeup of samples used across studies (e.g., students versus a mixture of racial groups), methods used in simulated trials to present race/ethnicity in cases (e.g., transcript versus video), and/or differences in the measures used (e.g., dichotomous versus continuous scales) (Devine & Caughlin, 2015; George & Martinez, 2002).

However, evidence suggests that racial bias in juror decision making is not as clear-cut as ingroup versus outgroup bias, but that racial stereotypes about who commits certain crimes can also influence decisions on guilt, blame and punishment (Gordon, 1993; Jones & Kaplan, 2003; Petsko & Bodenhausen, 2019; Smalarz et al., 2016). Race-crime congruency is the notion that certain racial groups are associated with certain crimes, with an alignment or match between the defendant's race and the crime considered stereotypical for the defendant's race, resulting in more guilty verdicts and sterner punishments than when there is a mismatch (Gordon et al., 1998; Sunnafrank & Fontes, 1983). Indeed, in a classic study by Sunnafrank and Fontes (1983), in which mock jurors were required to match photographs of five offenders (five Black and five White) with a list of ten crimes, they found mock jurors consistently associated Black offenders with blue-collar crimes (e.g., violent offences), and White offenders with White-collar crimes (e.g., embezzlement) (see appendix A for copy of Sunnafrank & Fonts list of race stereotypical time). Consequently, there will be circumstances when White rather than Black defendants are treated favourably and vice versa by jurors, because their racial ethnicity is congruent with the crime (Gordon, 1993; Jones & Kaplan, 2003). Several studies have supported this theory, suggesting that race-based crime stereotypes can impact on juror decisions (Levinson, 2007; Skorinko & Spellman, 2013).

Contrastingly, there has been a lack of research examining the possibility of crime stereotypes for Asian people. However, in the US at least, the concept of the 'model minority' has frequently been applied to East Asian people, based upon factors including their perceived high academic and economic successes, high family values and respect for people in authority (Lee, 2015). Such a halo effect

contrasts with the often-negative stereotypes about Black people, particularly in terms of Black people's associations with criminality (Eberhardt et al., 2004; Singh & Sprott, 2017). The misalignment with East Asian people and crime might contribute to the overall general leniency towards East Asian people in terms of conviction and sentencing decisions. However, this is only speculative, and more research needs to be done around crime-stereotypes for Asian people generally - including for the different Asian subgroups; and whether such crime-based stereotypes exist in England and Wales. A similar argument can be said for 'Black' and Black subgroups.

Interestingly, stereotypes about who commits certain crimes can impact on attributions on the causes of criminal behaviour and judgements of guilt and blame attributed to the victim (Miller & Turnball, 1986). If a defendant is considered typical for committing a given crime, then the perceiver is more likely to seek an internal cause for the criminal behaviour (Gordon, 1993). This suggests that if there is congruency between defendant race and the crime, then less blame would be attributed to the victim as compared to the situation when the defendant is accused of a race- incongruent crime. Indeed, a study by Bridges and Steen (1998) examined the influence of stereotypes on blame attribution and punishment recommendations in a sample of US probation officers. They found that probation officers were more inclined to attribute the crime committed by Black juveniles to internal causes (e.g., negative personality trait), and attribute the crime committed by White juveniles to external causes (e.g., dysfunctional family). Importantly, internal factors were also correlated with severer punishments and higher perceived recidivism rates. Similarly, other mock-juror studies have found that individuals whose crimes are linked to internal causes tend to be deemed more blameworthy and are recommended severer punishments than individuals whose crimes are considered linked to external causes (Lampinen, & Malesky, 2013; Peters et al., 2013; Templeton & Hartnagel, 2012). However, the race of the probation officer was not reported in the Bridges and Steen (1998) study, so it is unclear whether the race of criminal justice workers plays a role in their decision making. Further research is needed to ascertain whether race of criminal justice workers influences the attribution of blame and punishment recommendations. Furthermore, similar studies should be

conducted in England and Wales using a wider range of racial groups, to see whether similar results are found.

Limitations of juror and jury research

There are several features of previous juror and jury research that limit the representativeness and generalisability of the findings, including the overuse of students as participants, online recruitment platforms, and that most studies seem to be conducted in North America. There is also limited research using Asian defendants and victims, and the available literature is mostly focused on the East Asian population in the US, which is demographically different from the UK where people with a South Asian background constitute most of the Asian population (ONS, 2018). Furthermore, few studies have examined the effect of the criminal justice workers (CJWs) and type of criminal justice profession (CJPs) on decisions against Black and/or Asian versus White defendants and victims.

Thesis rationale

The current thesis focuses on exploring the views of CJWs/CJPs using the mock juror paradigm. The reason for focusing on CJWs/CJPs is that collectively they are involved in every stage of the criminal justice process, from the arrest and prosecution of suspects to the release of prisoners.

The fact that there is often an overrepresentation of Black and mixed-race people at in various stages of criminal justice system (e.g., stop and search: Lammy, 2017) has raised concerns of racial bias in the CJS (Lammy, 2017). It therefore seems necessary to explore whether the decisions of CJWs and CJPs, are influenced by race. This thesis then addresses the gap in knowledge around the possibility of racial bias in the decisions of criminal justice workers and criminal justice professions in England and Wales.

It is acknowledged that in England and Wales many criminal justice workers are prohibited from being jurors. Nevertheless, the mock juror paradigm was used in this thesis as an indirect way to assess whether race impacts on their decision making. Issues regarding socially desirable responding have already been outlined here (e.g., wanting to appear unbiased (Axt et al., 2016)) particularly in terms of it possibly invalidating study findings. Consequently, the use of mock-juror paradigm and case

vignettes offers a novel way of exploring CJWs'/ CJPs' views and decisions on sensitive topics including race, in a less intrusive way than direct discussions (Finch, 1987), which may minimise socially desirable responding and experimenter bias (Hughes & Huby, 2004; Schoenberg & Ravdal, 2000).

Importantly also, the criminal justice system in England and Wales differs considerably from the justice system in North America. The findings from such research can potentially inform on whether racial biases exist among criminal justice workers/professions, whether criminal justice workers/professions show racial biases against certain racial groups, whether racial biases differ across different criminal justice professions (CJPs), and whether ingroup bias exists amongst criminal justice workers and whether this might account for the racial disproportionality figures.

Overall, evidence from this research will add to the wider literature on understanding any racial differences in the CJS and helps to understand whether and how (if at all) race can influence the decisions of criminal justice workers. If the racial disproportionality is influenced by the biases held by criminal justice workers, then it should be known and addressed through changes in recruitment and training of criminal justice workers, or other interventions.

The aims of the thesis are fourfold:

1. To systematically review the literature on whether mock jurors treat Black defendants and victims differently from White defendants and victims on decisions of guilt, sentencing/ punishment, and blame.

   There has been a recent influx of studies within the area which warrants another systematic review, as attitudes and behaviours might have changed with time and since the last meta-analysis was completed. Furthermore, the focus here is on Black versus White defendants and victims, as not only do statistics indicate that Asian people are underrepresented in the England and Wales CJS, but a scoping exercise for the systematic review found a lack of studies that focused on decisions made by jurors regarding Asian defendants and victims.

2. A critique of the Balanced Inventory of Desirable Responding – 16 (BIDR-16) psychometric (Hart et al., 2015).

   The current research administers a self-report questionnaire to participants. The BIDR-16 was used to assess socially desirable responding amongst participants. If participants were providing socially desirable responses, then this could invalidate the study findings. Consequently, it is necessary to critique the BIDR-16 to better understand its validity and reliability, and so the validity and reliability of the results of this study.

3. To explore the influence of race (Black, White, Asian) on decisions on guilt, punishment and blame amongst criminal justice workers in England and Wales.

   There is a lack of research exploring the decisions of criminal justice workers in England and Wales, and from the scoping exercise as part of the systematic review, it was found there was a lack of research that has included decisions regarding Asian defendants and victims. The vital role that criminal justice workers place at all stages of the criminal justice decision-making process makes this an important area to research. The simulated method or mock juror paradigm was applied to observe the impact of defendant and victim race (Black, White, Asian) on criminal justice workers' decisions of guilt, punishment, and blame. The method was chosen over the archival method as it has the advantage of introducing experimental control, allowing for race of defendant and victim to be manipulated whilst keeping the facts of the crime constant, allowing for the assessment of the relevance of race on the dependent measures of verdict/guilt, punishment, and the attribution of blame.

   The criminal justice workers that formed part of this research included police, probation officers, youth justice/social workers, and prison officers.

4. To explore the influence of in-group bias in the attribution of guilt, sentencing and blame amongst White criminal justice workers in England and Wales.

   Note, alongside the overrepresentation of BAME people within the UK CJS, is the underrepresentation of BAME people within its workforce. A UK government report stated

that as at the end of March 2020, 7.3% of police officers in England and Wales who declared their ethnicity identified as BAME, with the figure decreasing across higher police ranks (Home Office Statistical Bulletin 20/20). Similarly, as at the end of March 2019, 7.1% of prison and probation staff (Ministry of Justice, 2019), 12% of magistrates and 6% of judges (UK Government, 2019) in England and Wales who declared their ethnicity identified as BAME. If there is racial bias against BAME people in the CJS, given the fact that most legal decisions made about BAME suspects, defendants and victims are made by White CJPs, it is important to assess whether such decisions are fair. If they are found to be racially biased, then this may go some way to explaining the racial disparity in the representation and treatment of BAME within the UK CJS.

Again, the simulated method was used due to the advantages outlined above. The criminal justice professions that formed part of this research included police, probation officers, youth justice/social workers, and prison officers.

Chapter 2: Do mock-jurors treat Black defendants and victims differently from White defendants and victims on decisions of guilt, blame and punishment? A systematic review.

As outlined in chapter one, the evidence on the influence of defendant and victim race on the verdict and punishment decisions of jurors and juries is mixed. There is evidence for and against defendant and victim race effecting jury and juror decisions (Ewanation & Maeder 2018; Maeder et al., 2015, study 2; Marcus-Newhall, Blake, & Baumann, 2002) and evidence for and against the race of defendant, victim, and juror interacting to affect juror decisions (FosterLee et al., 2006; Lee et al., 2019; Pfeifer & Ogloff, 2003; Saucier et al., 2010). To try to clarify the inconsistences in the literature, several meta-analyses have been conducted. The meta-analyses have focused on simulated research, perhaps because simulated studies use similar methods making it easy to compare findings across studies.

Previous meta-analysis

An early study Mazzella and Feingold (1994) examined 80 mock-juror studies and found no significant effect of race on juror verdict or sentencing decisions. However, the type of crime moderated racial bias on sentencing decisions, with longer sentences for Black defendants for negligent homicide offences, and for White defendants for fraud and embezzlement offences. However, this meta-analysis has since been criticised for not controlling for the influence of the mock-juror's race on decisions (Mitchell et al., 2005).

Other meta-analyses have found same-race leniency by mock jurors. Sweeney and Haney's (1992) meta-analysis of 14 early studies found a small effect of White jurors proposing shorter sentences for White defendants compared to Black defendants. The effect was strongest where the mock juror and victim race was controlled for, highlighting that to avoid potential distorted findings, these racial factors also need to be considered when examining the influence of defendant race on mock-juror decisions. Similarly, the meta-analysis by Mitchell et al. (2005) in 34 studies found racial bias to have a small significant effect on verdict, with both White and Black jurors having an increased likelihood to convict outgroup defendants. However, when verdict was analysed dichotomously rather than on a continuous scale, racial bias was non-significant. The latter is noteworthy as a dichotomous rating is arguably more ecologically valid to actual juror decisions than using a continuous scale.

A meta-analysis by Devine and Caughlin (2015) further found a significant small effect of outgroup bias by White mock jurors comparing White with Hispanic defendants, and a significant modest effect for outgroup bias by Black jurors against White defendants. They did not find ingroup bias by White mock jurors comparing White and Black defendants. Interestingly, ingroup bias was more noticeable amongst Black compared to White mock-jurors.

Current study rationale and aims.

Whilst some studies and meta-analyses have found racial biases in juror decisions, other studies have not. The rationale for the current systematic review is to try to help clarify this inconsistency. Previous meta-analyses are now dated, and attitudes and opinions might have changed over time. Similarly, newly conducted studies may now include those conducted outside of North America, enabling more generalisations to be made. Furthermore, the racial disproportionality of Black people particularly in the England and Wales CJS highlights the importance of understanding and identifying factors that might contribute to this occurrence.

A scoping exercise on the topic conducted prior to the systematic review found a lack of studies that assessed decisions for Asian defendants and victims. The lack of research might be led by the evidence that Asian people are generally underrepresented within most areas of the CJS (Lammy, 2017). Consequently, due to the lack of studies concerned with decisions involving Asian defendants and victims, this systematic review focused on the question: "*Do mock-jurors treat Black defendants and victims differently from White defendants and victims on decisions of guilt, blame and punishment?*" The question fits with the data that Black people are disproportionately overrepresented in the CJS. The review question sets out to uncover whether at least in terms of juror decisions, racial bias might account for the racial disproportionality figures in England and Wales.

Method

The present systematic review was undertaken according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (see Rethlefsen et al., 2021).

Literature search/ information sources. Articles for the systematic review were identified by (a) searching 11 electronic databases (1. PsychInfo; 2. Web of Science; 3. Criminal Justice Database; 4. APA Articles; 5. Social Science Database; 6. Psychology Database; 7.International Bibliography of the Social Sciences (IBSS); 8. ASSIA; 9. Criminology database; 10. Scopus; 11. ProQuest Dissertations & Theses A&I) (b) cross-referencing studies from the last meta-analysis (Devine & Caughlin, 2014); and (c) contacting researchers in the field to find out if they knew of any relevant unpublished studies or ongoing research in the field.

Inclusion criteria. The study population had to be mock juror and jury studies that used a mock juror/jury paradigm, as the primary studies in this thesis intended to put criminal justice workers (CJWs) in the role of mock juror to see if individually they show racial bias. Therefore, the intention was to discover what existing studies mock juror studies have found out about mock jurors' decisions on the subject. Studies included in the review also need to have used a cross sectional/survey study design; looked at decisions on Black defendants and/or victims compared to White defendants and/or victims; reported quantitative outcomes data on guilt, blame, punishment, and/or sentence (primarily as the review was interested in the decisions themselves, rather than the underlying reasons for those decisions); been written in English; and studies needed to be published between January 2015 and November 2021, which was chosen as the last meta-analysis on the topic was conducted by Devine and Caughlin in 2014. Therefore, the current systematic review can therefore be viewed as an update on an aspect of the Devine and Caughlin (2014) meta-analysis.

Exclusion criteria. The review excluded non-English written articles; exclusively qualitative research; studies that did not use a mock-juror paradigm; studies that used groups of participants other than mock-jurors; studies that did not identify defendants and victims to be Black or White; and studies

exclusively exploring other extra-legal factors, for example, socioeconomic status. In addition, archival studies were excluded.

Search strategy:

A range of key terms (see Table 1) were used to build the search strategy. Boolean operators "AND" and "OR" as well as truncation were employed to enhance the efficiency of the search.

Table 1

An example of the search strategy used to search the PsychInfo database

Database	Search String			
	Population	Intervention	Comparison	Outcome
PsychoInfo	("mock jur*" or "simulated jur*" or "mock trial*" or "simulated trial*" or "jur* selection")	(Black* or *afro Caribbean*" or Caribbean* or African* or "American African*" or "African American*" or BAME or BME or "ethnic* minorit*")	White*	(((race or racial* or ethnic*) adj4 (discriminat* or attitude* or bias* or prejudice*)) or (racial or racism or "rac* discriminat*" or "ethnic* discriminat*" or "rac* bias*" or "decision making" or "implicit bias" or "unconscious* bias*))

The search strategy was adapted if required for each given database search. Please see appendix B for a copy of the search strategy used for each database.

Email alerts for each database were set-up to keep abreast of any new research released. The last search was conducted on 30th November 2021.

Managing records. Citations for the systematic review were uploaded, stored, and organised in the reference management system Endnote.

Screening and selection of citations for review. There were two reviewers. The PRISMA-S flowchart/ checklist for screening studies was followed (Rethlefsen et al., 2021). After articles were identified through database searches and other sources, duplicates were removed. The first reviewer screened the titles and abstracts of each article to check if they met the eligibility criteria for inclusion in the

review, and those that did not were then excluded. The second reviewer also screened a sample of the citations at abstract and title level, and any discrepancies between reviewers' selection of citations were discussed until decisions to include/exclude citations were agreed upon with reasons for decisions given. The remaining included articles were then screened at full text by the first reviewer. Subsequently, the second reviewer screened a selection of citations to assess for eligibility. Articles that they agreed did not meet the inclusion criteria were excluded and reasons for exclusion provided. The resultant articles were selected for the review.

Data extraction. A data extraction form was developed using Microsoft Excel to record relevant study information for the review, including author(s); year of publication; study design; participant demographics and sample size; outcome measures used (e.g., verdict, punishment, blame); statistical analysis and results. See appendix C for a copy of the data extraction form.

Quality assessment. The Center for Evidence-Based Management's (CEBMa) critical appraisal tool for a cross-sectional study (July 2014) was used to critically appraise the studies for risk of bias (e.g., selection bias), generalisability, statistical issues (e.g., power), and other issues specific to the research question. The tool was adapted from Crombie, The Pocket Guide to Critical Appraisal; the critical appraisal approach used by the Oxford Centre for Evidence Medicine, checklists of the Dutch Cochrane Centre, BMJ editor's checklists and the checklists of the EPPI Centre. Although alternative appraisal tools for cross-sectional studies could have been used (e.g., AXIS, CASP), the CEBMa was chosen for its conciseness and simplicity of use. Furthermore, the CEBMa tool covers some questions/ domains also used in other longer cross-sectional critical appraisal tools (e.g., related to appraising the method and results section).

Narrative synthesis. The narrative synthesis was done to provide a written summary of the findings from all the studies included in the review. The narrative synthesis included a summary of the study type, participant characteristics, outcome measures used, quality of the study, a discussion of the

heterogeneity across studies, and a description of the pattern of the data (e.g., trends of findings) across the included studies. The studies were cited in text based on the study numbers provided in table 2 in the results section overpage. Table 2 provides a summary of the study characteristics of the included studies that formed part of this review.

Results

Table 2

Study characteristics

Author	Population	Offence	Presentation of Scenario	Race manipulation	Study design
1. Gamblin & Kehn (2021)	207 undergraduate White mock jurors. Age ranged 18–45 years old (19.57± 1.8 years); 66.7% female.	Assault	Trial transcript	Defendant (Black/White)	2 (defendant race: White vs. Black) X 4 (race salience: Opening and Closing vs. Opening Only vs. Closing Only vs. Not Salient)
2. Silberkleit (2021)	669 undergraduates. Age range 18-43 years old (19.45± 2.1 years). No further breakdown of participant gender and ethnicities was provided.	Sexual offence (child sex trafficking)	Trial transcript & photos of defendant and victim	Defendant (Black/White) & victim (Black/White)	2 (defendant race: White vs. Black) X 2 (IAT and mock trial order) X 2 (victim race)
3. Ellis (2020)	1011 participants were recruited via Amazon Mechanical Turk (MTURK). Age ranged from 18-79 years old (*M* = 40.11, *SD* = 12.99). 53.5% female, 45.7% male, 0.8% nonbinary/genderqueer/other. Participants' ethnicity: European American/White (72.8%); Black (9.5%), Asian/Asian American/Pacific Islander (5.5%) as, as Hispanic/Latin American (5.1%), as Biracial/Multicultural (5.2%), Native American/American	Second degree Burglary & third-degree Arson	Trial transcript & photo of defendant	Defendant (Black/White)	2 (defendant race: Black/White) X 2 (defendant age: 16 years old/18 years old) X 2 (crime: second degree burglary/third degree arson)

	Indian/ Indian Indigenous (1.2%), Middle Eastern and other (0.7%).				
4. Maeder & McManus (2020)	American sample: 191 participants recruited by MTURK. Mean age 38.42±11.9 years. 34 (37%) men, 58 (62%) women, and one (1%) transgendered person. 83% White, 7% Black, 2% Asian, 1% LatinX, 2% American Indian, 5% other ethnicities. Canadian sample: 105 participants recruited via Qualtrics Panel. Mean age 52.7 years (*SD* = 15.71). 87% White, 6% Asian, 2% Black (2%), 2% Latinx, and 4% other ethnicities.	Assault	Trial transcript	Defendant (Black/White & Aboriginal-Canadian/Native-American)	3 (defendant race: White, Black, Aboriginal Canadian/Native American) X 2 (country: Canada, United States)
5. Maeder, Yamamoto, & McLaughlin (2020)	216 U.S. jury eligible community members, recruited via MTURK. Age: 36.7± 12.6 years old. 51.9% women, 46.8% men, 1.4% transgender individuals. 82.9% White, 5.1% Black, 4.2% Asian, 2.8% Hispanic/Latino, 1.4% Native American, 3.4% other.	Second degree murder	Trial transcript	Defendant (Black/White)	2 (defendant race: Black/White) X mental disorder (schizophrenia/depression)
6. Tang, et al. (2020)	151 undergraduate students. Age range 18–51 years (20.11± 0.29 years).	First degree murder	Trial transcript	Defendant (Black/White)	2 (juvenile defendant race: Black, White) X 2 (intellectual disability: disabled, nondisabled)

	38 males and 113 females. 68.2% Caucasian, 9.3% Black, 7.9% Asian, 6% Hispanic, 8.6% Other.				
7. Bell (2019)	1000 university students. Age range 18 -72 years (Median 21 years). 23% were males, and more than 2% belong to the other or the transgender category. 61.9% White; 25.5% African Americans; 5.1% Hispanic; 1.7% Asian, 5.4% other	Murder	Written vignette & photo of defendant	Defendant (Black/White) & Victim (Black/White)	2 (female defendant race: Black/White) X 2 (victim race: Black/White) X2 (female defendant: attractive/unattractive)) X 2 (victim gender: male/female)
8. Burke (2019)	163 all-White Jury-eligible US citizens recruited via MTURK. Age range 19–72 years (37.79± 11.7 years). 48.5% of participants were women.	Residential Burglary	Video trial summary	Defendant (Black/White)	2(Defendant race: Black, White) X 3(Priming condition: no harm control, harm control, collective guilt) design X 2 (White identification: high, low)
9. Call (2019)	261 jury-eligible participants recruited via MTURK. Age range: 18-73 years (37.89 ±12.24 years). 137 women, 123 men. 70% White, 10% Black.	Sexual offence	Trial transcript & photo of defendant and victim	Defendant (Black/White) & victim (Black/White)	2 (defendant race: Black vs. White) X 2 child race (Black vs. White) X 2 child age (12 years vs. 17 years)
10. Kurinec & Weaver, III (2019)	548 participants recruited via MTURK. Mean age = 37.11 years. 52.3% were male. Participants' ethnicities were: White (78.9%); Black (9.9%), Hispanic or Latino/a	Burglary	Trial case materials & audio recording	Defendant (Black/White)	2 (Race: White or Black) X 2 (Dialect: General American English (GAE) or African American Vernacular English (AAVE)

	(1.0%), and Asian American (5.8%).				X 2 (Case: Ambiguous or Pro-Prosecution)
11. Maeder & Yamamoto (2019a)	558 jury-eligible community members recruited via Amazon MTURK. Age range: 18-74 years old (30.0±10.6 years). 288 (52%) men, 264 (47%) women, and 5 (1%) identifying as another gender. Participants' ethnicities were: 72% White, 12% Asian, 5% Black, 2% Aboriginal Canadian, 1% Latinx (1%), and 8% other.	Robbery	Trial transcript & photos of defendant and victim	Defendant race: White, Black, Aboriginal Canadian victim race: White, Black, Aboriginal Canadian	3 (defendant race: White, Black, Aboriginal Canadian) X 3 (victim race: White, Black, Aboriginal Canadian) X 2 (manipulated race salience: salient, not salient)
12. Maeder & Yamamoto (2019b)	276 jury-eligible community members recruited through a Qualtrics Panel. Mean age of 39.3 (*SD* = 15.0). 181 (66%) were women and 91 (33%) men, with 3 (1%) identifying as another gender. 33% Black, 35% White, and 33% Indigenous.	Impaired driving	Trial transcript	Defendant (Black/White/ indigenous)	The defendant's race varied (Black, Indigenous, White).
13. Nesbit (2019)	Study1: 182 participants recruited via Amazon MTURK. Age range: 18 -65 years (31.97±7.85 years). 53.3% male. Participants' ethnicities were: White (50.5%) and Black (49.5%).	Armed robbery	Trial transcript & photo of defendant and eyewitness	Defendant (Black/White/ Hispanic) eyewitness (Black/White/ Hispanic)	Study 1: 2 (defendant race: Black/White) X 2 (eyewitness race: Black/White) X 2 (mock juror race: Black/White) Study 2: 3 (defendant race: Black/White/Hispanic) X 3

	Study 2: 214 participants 18-65 years old (31.66±8.51 years). White (*n* = 75), Black (*n* = 73), and Hispanic (*n* = 66).				(eyewitness race: Black/White/ Hispanic) X 3 (mock juror race: Black/White/ Hispanic)
14. Maeder & Ewanation (2018)	385 American jury-eligible community members recruited via Amazon MTURK. Age range: 18-73 years old (44.39±14.38). 50% White and 50% Black	Second degree murder	Trial transcript	Defendant (Black/White/ indigenous)	3 (defendant race: White, Black, Indigenous) X 3 (juror race: White, Black, Indigenous)
15. McManus, Maeder & Yamamoto (2018)	201 jury-eligible Canadian community members recruited using a Qualtrics Panel. Mean age of 48 (*SD* = 15.82). 86 (41%) men and 124 (59%) women. 178 (84.8% White, 6.2% Asian, 2.4% Black, 5 (2.4%) 1.9% Aboriginal Canadian, 4.8% other.	Dangerous operation of a motor vehicle (DOM) and impaired driving causing bodily harm	Trial transcript & photo of defendant	Defendant (Black/White/ indigenous)	2 (Defendant race: Black, Indigenous) X 3 (Race salience: Specific, General, Neutral) A White/race neutral control condition was included for comparison with the minority defendant/ race neutral conditions.
16. Armenta (2017)	375 undergraduate students. Age range: 18–31 years (19.1±2.23 years). Most participants were 18 years of age (57.6%). 70% were female; 30% male. 38.4% Hispanic, 32.3% White, 17.3% Asian American; 2.7% African American, 0.5% Native American, and other 8.8%.	Murder (filicide: mother against child)	Trial transcript & photo of defendant	Defendant (Black/White/ Hispanic)	3 (Ethnicity: Black, White, or Hispanic) X 2 (socioeconomic status: low or high) X 2 (defendant attractiveness: attractive or non-attractive)

17. DeCesare (2017)	108 undergraduates aged 18 to 38 years (21.17±4.36 years). 75.2% were female and 24.8% male. 23.9% Caucasian, 11.9% African American, 51.4% Hispanic and 12.8% Other.	Robbery & Murder	Trial transcript	Defendant (Black/White/ Hispanic)	3 (defendant race/ethnicity: White, Black, or Hispanic) X 2 (verdict condition: guilty/not guilty) or X 3 (verdict options: guilty of second-degree murder (primary charge), guilty of voluntary manslaughter (lesser charge) or not guilty).
18. Leippe, et al. (2017)	237 college students. 42 Black, 134 Hispanic, and 61 non-Hispanic White. No gender or age breakdown provided.	Murder & Carjacking	Trial transcript & photos of the defendant	Defendant (Black/White/ Hispanic)	3 (defendant race/ethnicity: Black, Hispanic, or White) X 2 (type of crime: murder or theft) X 2 (mortality salience: present or absent)
19. Pica et al. (2017)	444 jury-age eligible psychology students. Age range: 18 to 55 years (20.61±4.51 years). 77% female. 72.7% White, 14.6% Asians, 1.1% Latino/Latinas,0.2% Aboriginals, 5.2% Black, and 6.1% other.	Assault	Trial transcript	Defendant (Black/White/Aboriginal-Canadian)	3 (Defendant race: White, Black, or Aboriginal Canadian) X 2 (Defendant's developmental age: 14 or 24 years old) x 2 (defendant's chronological age: 14 or 24 years old)
20. Yamamoto	97 jury eligible participants recruited via Amazon MTURK.	Looting	Trial transcript &	Defendant (Black/White) & mock juror (Black/White)	Mock juror race (Black/White) X defendant race (Black/White)

& Maeder (2017)	Age range: 18-67 years old (34.6±11.9 years). 58.2% White and 40.8% Black/African American.		photo of defendant		
21. David (2016)	194 introductory psychology students. mean age: 19 years. 29% were males and 71% females. 52.5% Caucasian, 16.5% Asian and 3% African American.	carjacking	Trial transcript	Defendant (Black/White)	2 (defendant race: Black/White) X 2 (presence or absence of pro-prosecution metaphor ("beast" metaphor) or pro-defence metaphor ("officer was a bully") X 2 (presence or absence of a defence metaphor ("bully metaphor")
22. Maeder et al. (2016)	397 jury-eligible university students. Mean age of 20.8 years (*SD* = 5.1). 72% women, 28% men. 63.7% White,7.3% Asian, 7% Black, 3.8% Latino/a, 1.3% Aboriginal Canadian, and 16.9% other.	dangerous operation of a motor vehicle while intoxicated/ fraud	Trial transcript	Defendant (Black/White/ Aboriginal-Canadian)	3 (Defendant Race [White, Black, Aboriginal Canadian] X 3 (Crime Type: fraud, theft of a motor vehicle, Dangerous Operation of a Motor vehicle)
23. Dodson (2015)	354 juror eligible participants recruited via a Qualtrics Panel. Mean age: 47.6 (*SD* = 13.60) years old. 61% (*n* = 217) were female, 82% White.	First degree murder & armed robbery	Trial Transcript	Defendant (Black/White/ Hispanic)	3 (race: White, Black, or Hispanic) X 2 (crime: Armed Robbery or Armed Robbery and First-Degree Murder) X 2 (sex: male vs. female).

24. Glaser et al. (2015)	276 American adults via Time-Sharing Experiments in the Social Sciences (TESS). Median age 46 years. 50% women; 84.8% White, 6.2% Hispanic, and 4.7% Black.	Triple Murder	Trial transcript	Defendant (Black/White)	2 (Defendant Race: Black; White) X 2 (Maximum Sentence: Life Without the Possibility of Parole; Death Penalty)
25. Kluwe (2015)	299 White participants recruited via Amazon MTURK. Age range:18-80 years (35.21±12.04 years). 54.5% were female.	Grand-theft auto/ embezzlement	Trial transcript & photo of defendant	Defendant (Black/White)	2 (defendant race: Black/White) X 2 (type of crime: grand theft auto/ embezzlement) design
26. Maeder, Yamamoto, & McManus (2015)	Study 1: 137 jury-eligible students (34 men and 103 women) recruited via an online participant pool. Mean age of the participants was 20.8 (*SD* = 4.0). 67.9% were White, 9.5% were Black, 8.8% Asian, 8.8% Middle Eastern, 2.2% East Indian, 1.4% Latino/a, and 1.4% Aboriginal Canadian. Study 2: 120 jury-eligible students (33 men and 87 women) recruited via an online participant pool. The mean age of the participants was 20.9 (*SD* = 4.7). 62.5% were White, 10% were Black, 5.8% Asian, 5.8% Middle Eastern, 1.7% Hispanic,	Robbery & theft of a motor vehicle	Trial transcript & photograph of the defendant	Defendant (Black/White/ Aboriginal-Canadian)	2 (Defendant Race: Black, Aboriginal Canadian) X 2 (Race Salience: present, absent).

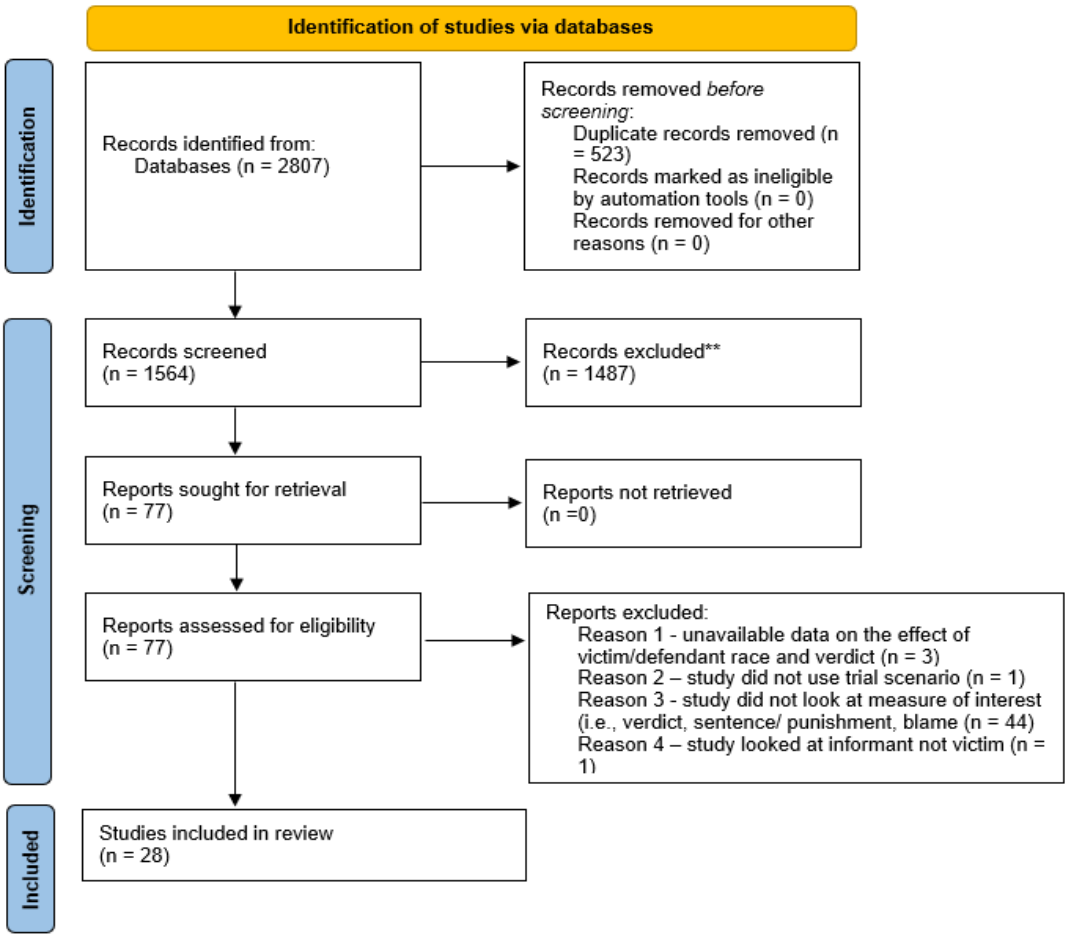
	1.7% East Indian, and 1.7% Aboriginal Canadian (10.8% identified as another race).				
27. Maeder, Yamamoto & Saliba (2015)	125 White student participants. Age range: 17–50 years old (20.5±3.71). 59 men and 66 women).	Sexual Assault	Trial transcript & photograph of defendant	Defendant (Black/White/ Aboriginal-Canadian)	3 (defendant race: White, Black, Aboriginal) X 2 (victim attractiveness: attractive, unattractive)
28. Piatak (2015)	279 college students from a mid-sized South-eastern university. No details of age provided. White (81.9%), Black (7.2%) and other races (10.8 %). Female: 58.3%; Male 41.7%.	Sexual offence	Trial transcript	Victim (Black/White/ Hispanic)	3 (victim race: Black, White, Hispanic) X 2 (victim gender: male/female).

Study selection

As depicted in figure 1, 2807 publications were identified during initial search through 11 databases and 1564 records remained after deduplication. Seventy-seven publications remained after screening the titles and abstracts. Twenty-eight publications remained after reading full texts and were included in the systematic review (also see table 2).

Figure 1

Steps of the PRISMA process and the number of citations excluded at each stage.



From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

Study characteristics

Study sample. A total sample of 9963 mock jurors (participants) were included in this review, with a sample size range of 97 ([20]) to 1011 mock jurors ([3]). Nineteen of the reviewed studies were conducted in the US, seven in Canada, and two jointly in the US and Canada.

Demographically, the pooled ages of mock jurors ranged from 18-80 years old. Most studies recruited an even split of male and female participants, with nine studies reporting to have samples with less than 40% male participants ([4,6,7,12,16,17,19,21,26: study2]). Nineteen studies used predominately female participants.

Regarding racial breakdown, four studies recruited only White participants ([1,8,25,26]); most studies had majority White participants, with 15 recruiting less than 10% Black participants ([3,4,5,6,9,10,11,15,16,19,21,22,24,26: study1, 28]).

Twenty studies used a mixture of Black, White and another race as mock jurors; three studies used Black and White mock jurors ([9,14,20]); three studies only White mock jurors ([8, 25, 27]). One study only reported the percentage of White mock jurors used and did not provide a racial breakdown of other mock jurors ([23]); and one study did not report the racial composition of the sample used ([2]). Of the studies that included Black and White mock jurors, 15 reported that at least 70% were White, and 13 studies reported that less than 10% of the mock jurors used were Black.

Thirteen studies recruited students as participants ([1,2,6,7,16,17,18,19,21,22,26,27,28]). Sixteen studies recruited participants via online recruiting platforms (i.e., Amazon MTurk; Qualtrics Panel; Time-Sharing Experiments in the Social Sciences (TESS) ([3,4,5,8,9,10,11,12,13,14,15,20,23,24,25,26]).

Study design and race manipulation. All studies within the review used the simulated mock-juror method described earlier (Landy & Aronson, 1969), whereby the dependent variable of interest was manipulated (e.g., race of defendant/victim), whilst other facts of the case held constant (e.g., the evidence).

The studies in this review manipulated defendant and victim using either a X2 (Black/White) or X3 (Black/White/another race) design. Twenty studies manipulated only the defendant race; one just the victim race ([28]); and four both defendant and victim race ([2,7,9,12]). In conjunction with manipulating the defendant or victim race, two studies also manipulated eyewitness and mock juror race ([13, 20]).

Other (extra-legal) factors studied. Other variables manipulated in conjunction with defendant and victim race, included race salience (*n* = 5); type of offence (*n* = 6); social economic status (*n* = 1); defendant/victim physical attractiveness (*n* = 3); defendant/victim gender (*n* = 3); defendant's mental disorder (*n* = 1); nationality of mock juror (*n* = 1); age of defendant/victim (*n* = 3); defendant's intellectual disability (*n* = 2); and defendant's dialect (*n* = 1).

Crimes used within trial scenarios. A total of eleven categories of crime were used in the case trials presented to mock jurors: assault (*n* = 3), adult sexual offending (*n* = 3), child sexual offending (*n* = 1), murder (*n* = 9), burglary (*n*= 3), robbery (*n* = 5), carjacking (*n* = 2), driving offences (*n* = 5), looting (*n* = 1), embezzlement (*n* = 2) and arson (*n* = 1). Twenty studies had only one criminal charge depicted in the case trial; whilst seven studies depicted two of more criminal charges in the case trial.

Presentation of scenario. The race of the defendant and victim was manipulated through either (or all) of the following: trial transcript, audio recording, video recording, a photograph of the defendant/victim. Thirteen studies manipulated race using a transcript ([1,4,5,6,12,14,17,19,21,22,23,24,28]); eight studies used a transcript plus a photograph of the defendant ([7,15,16,18,20,25,26,27]); and three studies used a transcript plus a photograph of the defendant and victim ([2,9,11]). Additionally, one manipulated the race of the defendant and victim via a video recording ([8]); and one study via a transcript plus photographs of the defendant and witness ([13]).

Outcome measures. Outcome measures examined mock-jurors' decisions on guilt, blame and punishment/sentencing.

Verdict. Guilt was measured using binary (e.g., guilty/not guilty), categorical and continuous scales. Twenty-five studies measured guilt: 17 used a binary scale; seven a continuous and binary scales ([1,8,9,19,21,25,27]); and one a binary plus another categorical measurement ([17]). No study used just a continuous scale to measure guilt.

Punishment/sentencing. Punishment/sentencing was measured using binary, categorical, and continuous rating scales. Eleven studies measured punishment/sentencing decisions: seven used a continuous scale ([1,3,9,11,12,15,20,27]); one used a binary scale ([7]); one used a categorical scale (e.g., no sentence/prison with possibility of parole after 20 years) ([16]); and one used both a continuous and categorical scale ([13]).

Blame. Ten studies measured mock-jurors' decisions on blame using a continuous scale ([2,4,8,9,15,16,22,23,27,28]).

Study results. Table 3 overpage summarises the design, outcome measures and findings from the studies in this review.

Table 3

SR study results - summarises the design, outcome measures and findings from the studies in this review.

Author	Offence	Outcome Measures	Findings
1.Gamblin & Kehn (2021)	Assault	Verdict: guilty/ not guilty & guilty scale (1 to 9) Recommended sentence (length in years and months).	Guilt: main effect of defendant race (p = .009). Black defendants (M = 4.03, SD = 1.46) were rated as less guilty than White defendants (M = 4.84, SD = 1.39). Sentencing recommendations: non-significant main effect and interaction effect of defendant race
2.Silberkleit (2021)	Sexual offence (child sex trafficking)	Verdict: guilty/ not guilty Culpability (sympathy) towards victims	Guilt: Defendant race a significant predictor (OR: 0.42, p = .001). Mock jurors significantly more likely to find White defendants guilty (58.5%) than to find Black defendants guilty (40.7%). Victim blame: Significant effect of Defendant race (p = .002). More culpability attributed to the victim of a Black defendant (M = 2.95, SD = 1.07) than a White defendant (M = 2.57, SD = 0.98). Defendant blame: significant effect of defendant race (p <.001). Higher culpability for a White defendant (M = 4.27, SD = 1.18) than a Black defendant (M =3.58, SD = 1.27). The interactions of defendant and victim race on verdict, defendant culpability and victim culpability were non-significant.
3.Ellis (2020)	Second degree Burglary & third-degree Arson	Verdict: guilty/ not guilty recommended sentencing (sentence length from 1 to 15 years)	Guilt: significant interaction of defendant race and crime type. For Black defendants, the association between crime type and verdict was non-significant. for White defendants this association was significant (p < .001). The percentage of guilty verdicts rendered for a White defendant was greater for arson charge (63.7%) than burglary (36.3%). Sentencing: Main effect of defendant race and interaction of race and crime type non-significant.

4. Maeder & McManus (2020)	Assault	Verdict: Guilty/not guilty. Blame (culpability) attribution (locus of control, stability, and controllability on 9-point scales)	Guilt: main effect of defendant race and interaction of race and country of study nonsignificant. Culpability: This analysis revealed a significant main effect of defendant race (F = (2, 180) = 3.75, p = .02). Participants provided more external attributions for the Black defendant (M = 4.84, SD 1.79) than for the White defendant (M = 3.86, SD = 2.13; p <. 001).
5. Maeder, Yamamoto, & McLaughlin (2020)	Second degree murder	Verdict: guilty/ not guilty	Guilt: main effects of defendant race nonsignificant. Significant interaction of race and mental disorder a Black defendant with schizophrenia was six times more likely to be found guilty than a Black defendant with depression (OR = 6.12, p = .013). The association of mental disorder with verdict was non-significant for a White defendant.
6. Tang et. al., (2020)	First degree murder	Verdict: Guilty/ not guilty	Guilt: Significant Main effect of defendant race (OR = 0.42, p = 0.01). Black defendants were less likely to receive a guilty verdict compared to White defendants (31% vs 51%). Significant interaction of defendant race and intellectual disability (OR = 0.23, p = 0.03). Mock jurors were less likely to find nondisabled Black defendants guilty than nondisabled White defendants (24% vs 61%). For disabled defendants, guilty verdicts were comparable for Black and White defendants.
7. Bell (2019)	Murder	Sentence: death/ life imprisonment without the possibility of parole (LWOP).	Sentence: Respondents were less likely to select the death penalty for African American offenders than White offenders (OR = 0.66, p <0.05), but more likely to select the death penalty in vignettes with African American victims than White victims (OR = 1.48, p <0.05).
8. Burke (2019)	Residential Burglary	Verdict: Guilty/not guilty Perceived likelihood of crime (11-point scale ranging from 0% (Not at all likely) to 100% (Extremely likely). Blame attributions (6-point scale).	Verdict: Significant main effect of defendant race (p = .02). Jurors were significantly more lenient toward the Black (M = 10.78, SD = 7.93) than the White (M = 13.44, SD = 6.95) defendant on the degree of guilt. Significant main effect of defendant race (p = .04) and White identification (p = .05) on the perceived likelihood of guilt. Jurors were significantly less likely to think that the Black (M = 5.55, SD =3.02) versus White (M = 6.45, SD = 2.37) defendant committed the crime.

			Blame: Significant effect of defendant race on internal attribution (p = .04). Jurors were significantly less likely to make internal attributions for the Black (*M* = 2.22, SD = 1.60) than the White (*M* =2.70, *SD* = 1.45) defendant.
9. Call (2019)	Sexual offence	Verdict: guilty or not guilty guilt rating (7 point) Culpability judgements (5point scale and categorical scale: how much the student/teacher was (a) responsible for, (b) wanted, and (c) caused the sexual contact) Sentencing recommendation: (from 0 to 15 years with an open text box option to extend it past 15 years).	Verdict: No significant main effects of defendant and child race. Guilt rating: Interaction of defendant race and child race was non-significant Significant interaction of defendant race, child race and child age (*p* = .031, ηp^2 = .02). In cases that involved a 17-year-old Black child, mock jurors attributed less guilt to the Black defendant than the White defendant, (*p* = .049, *d* = .49) when the case involved a 17-year-old White child, mock jurors' attributions of defendant guilt were similar whether the defendant was Black or White Sentencing: No significant main effects of defendant and child race Significant interaction of defendant race and child age (*p* = .006, ηp^2 = .08). Defendants received a greater sentence when the case involved a 12-year-old complainant versus a 17-year-old complainant. culpability ratings: main effects of defendant and child race nonsignificant. No significant interactions
11. Maeder &. Yamamoto (2019a)	Robbery	Verdict: Guilty/ not guilty Sentence length (scaled from 1 to 10)	Verdict: Main effect of defendant race or its interactions with juror race and stereotype nonsignificant Sentence length: significant main effect of defendant race (*p* = .002). Black defendant (*M* = 4.15, *SD* = 2.57) received a significantly shorter sentence length than the White defendant (*M* = 5.04, *SD* = 2.56)
12. Maeder & Yamamoto (2019b)	Impaired driving	Verdict: Guilty/ not guilty Sentence length (scaled from 1-10 i.e., no punishment to maximum punishment allowed by law)	Guilt: Main effect of defendant race on DOM or impaired driving nonsignificant DOM: A significant main effect of defendant race, *F* (2, 159) = 2.98, *p* = .05, η^2 = .04. The White defendant (*M* = 5.21, *SD* = 2.03) received a significantly higher sentence recommendation than did the Black defendant (*M* = 4.43, SD = 2.27), *p* = .01 Impaired Driving: A significant main effect of defendant race, *F* (2, 165) = 6.30, *p* = .002, η^2 = .07. The Black defendant (*M* = 4.15, *SD* = 2.57) received a

			significantly lower sentence recommendation than the White defendant (*M* = 5.04, SD = 2.56, *p* = .001)
13. Nesbit (2019)	Armed robbery	Guilt ratings: guilty/not guilty Sentence recommendations: life imprisonment (dichotomous) imprisonment (between 5-99 years) monetary fine (between $0 and $10,000)	Guilt: Main effects or factorial interactions of defendant, mock juror, and eyewitness race/ethnicity nonsignificant. Sentence: main effects or factorial interactions of defendant, mock juror, and eyewitness race/ethnicity nonsignificant.
14. Maeder & Ewanation (2018)	Second degree murder	Verdict: Guilty/not guilty	Same race identification condition: main effect of defendant race or juror race or their interaction on the verdict was non-significant. Cross race identification condition: significant interaction effect of defendant and juror race (*p* = .004). Black jurors were more likely to convict White defendants (67%) than Black defendants (41%). White jurors, the effect of defendant race on verdict was nonsignificant
15. McManus, Maeder & Yamamoto (2018)	Dangerous operation of a motor vehicle (DOM) and impaired driving causing bodily harm	Verdict: Guilty/not guilty Sentencing recommendation (scale from 1 to 9 (1 (no punishment) to 9 (maximum punishment allowed by law)). Attributions of the cause of the defendant's criminal behaviour locus of causality and stability	Guilt: effects of defendant race on verdict judgements nonsignificant. Sentencing: significant interaction of defendant race and article type (*p* =.005). For the Black defendant, races specific article yielded the harshest recommended sentence (*M* =7.89, *SD* = 1.94), compared to the race neutral (*M* 5.17, *SD* 2.15) and general race articles (*M* = 4.67, *SD* = 1.45) for the Indigenous defendant, significantly harsher sentences in the general race (*M* = 6.15, *SD* = 2.58) and race-specific conditions (*M* = 5.92, *SD* = 1.55) Blame: Significant main effect of defendant race, (*p* = .02). More external attributions for the Black defendant (*M* = 4.94, *SD* = 1.79) than for the White defendant (*M* = 3.86, *SD* = 2.13).
16. Armenta (2017)	Murder (filicide: mother	Verdict: guilty/ not guilty Sentence: length of sentence (no sentence, prison with the	Guilt: effects of defendant race, SES, and physical attractiveness nonsignificant.

	against child)	possibility of parole after 20 years, life in prison without the possibility of parole), Defendant culpability ratings (7-point scale).	Sentence: main effect of defendant race nonsignificant but significant interaction of defendant race, SES, and physical attractiveness (p = .001) Black and Hispanic defendants with low SES that were considered unattractive were given lengthier sentencing (life in prison without the possibility of parole) than all other conditions. Blame: Significant main effect of defendant race (p = .016). Black (M = 6.14) and Hispanic (M = 6.18) defendants were more likely to be considered responsible for the crime than White defendants (M = 5.78).
17. DeCesare (2017)	Robbery & Murder	Verdict preference: guilty of second-degree murder (primary charge), guilty of voluntary manslaughter (lesser charge) or not guilty defendant's likelihood of guilt.	Guilt: nonsignificant difference in ingroup and outgroup preference for a guilty verdict on both second-degree murder and voluntary manslaughter.
18. Leippe et al. (2017)	Murder & Carjacking	Verdict: Guilty/not guilty	Guilt: Significant interaction of defendant and mock juror race, indicating outgroup bias. Black mock jurors significantly (p = .041) more likely to prefer a guilty verdict for a Hispanic (69.2%) rather than a Black defendant (27.3%)
19. Pica, et al. (2017)	Assault	Verdict: Guilty/not guilty and a continuous scale (1 = not guilty;100 = guilty).	Guilt: Significant main effect of defendant race on mock jurors' dichotomous verdicts, OR = .61, 95% CI [.37, .99]. There were fewer guilty verdicts for the Black defendant (59%) compared to the White defendant (.70%). There was nonsignificant interaction of defendant race, developmental age, and chronological age (p = .006) on the verdict.
20. Yamamoto & Maeder (2017)	Looting	Verdict: Guilty/not guilty Sentence length (1 to 10 years)	Guilt: main effect of defendant race or its interactions with juror race and stereotype nonsignificant Sentence: a significant main effect of defendant race (p = .002). Black defendant (M = 4.15, SD = 2.57) received a lower sentence recommendation than the White defendant (M = 5.04, SD = 2.56)

21. David (2016)	carjacking	Verdict: Guilty/not guilty Probability of guilt rating (scale 1–10).	Guilt: Main effect of defendants' race or its factorial interactions with the defence or prosecutor metaphor on the likelihood of guilt and verdict were non-significant
22. Maeder et al., (2016)	dangerous operation of a motor vehicle while intoxicated/ fraud	Verdict: Guilty/not guilty Blame rating (from 1 to 9)	Guilt: crime type and defendant race interacted significantly on the verdict (p = .03). Black defendants were less likely than White defendants to be found guilty of auto theft (34.1% vs 48.8%) and Dangerous Operation of a Motor vehicle (36.7% vs 50%). Within the Black race, defendants received a greater proportion of guilty verdicts for auto theft trial compared to Dangerous Operation of a Motor vehicle. Guilty verdicts did not significantly differ for auto theft and fraud trial. Blame: main effect of defendant race non-significant. A significant interaction of defendant race and crime type (p = .02). In White defendants: blame ratings significantly higher (p = .001) in the fraud trial (M = 7.2, SD = 1.1) than in the Dangerous Operation of a Motor vehicle trial (M = 4.1, SD = 1.2) and in the auto theft trial (M = 7.1, SD = 1.3) than the Dangerous Operation of a Motor vehicle trial (p = .001). In Black defendants: highest blame ratings for auto theft trial (M = 7.4, SD = 1.1) than the fraud trial (p = .01, M = 6.2, SD = 1.5) and Dangerous Operation of a Motor vehicle trial (p = .001, M = 4.4, SD = 1.8). Also, significantly more blame attributions (p = .001) for the fraud trial than Dangerous Operation of a Motor vehicle trial. Results demonstrated only modest evidence of a congruency effect for these crimes.
23. Dodson (2015)	First degree murder & armed robbery	Culpability (scale 1-5).	Blame: a significant difference (p <0.05) in perceived culpability for Hispanic females (M = 4.39, SD = 0.90) than to non-Hispanic females (M = 4.09, SD = 1.01). However, the differences in perceived culpability were non-significant for Black vs non-Black or White vs non-White male or female defendants.

24. Glaser et al. (2015)	Triple Murder	Verdict: Guilty/not guilty	Guilt: a significant main effect of defendant race (p = .021) a greater tendency to convict ostensibly Black (73.9%) than White (60.9%) defendants.
25. Kluwe (2015)	Grand theft auto/ embezzlement	Verdict: Guilty/not guilty Probability of defendant committed a crime (scale -7 to 7).	Guilt: Main effect of defendant's race or its factorial interactions with crime type, stereotype, suppression Instructions or the need for cognition on verdict or the probability of guilt were non-significant. No race-crime congruence effects emerged.
26. Maeder, Yamamoto, & McManus (2015)	Robbery & theft of a motor vehicle	Verdict: Guilty/not guilty	Study 1: There was no significant difference in verdict decision between the White defendant and Black defendant conditions, B = .65, SE = .57, p = .25, odds ratio = .19. Study 2: Participants were more likely to vote guilty when the defendant was Black as compared with when the individual was White. In the trials involving a Black defendant, 35.7% (n = 15) voted not guilty, as compared with 64.3% (n = 27) who voted guilty.
27. Maeder, Yamamoto & Saliba (2015)	Sexual Assault	Verdict: Guilty/not guilty Verdict certainty scale (1 to 9) Sentence continuous (up to 10 years); Blame: defendant and victim responsibility	Guilt: Main effect of defendant race on verdict, verdict certainty and defendant responsibility rating nonsignificant. Black and Aboriginal Canadian defendants would be treated similarly to the White defendant. Sentencing: a significant main effect of defendant race (p < .05). Participants who voted guilty assigned a harsher sentence to the Aboriginal Canadian defendant (M = 4.91) than to the White (M = 3.13) and Black (M = 3.35) defendant. Blame: a significant interaction between victim attractiveness and defendant race (p < .05).

			When the defendant was White, attractive victims (*M* = 4.38) received higher responsibility ratings than did unattractive victims (*M* = 3.36). This effect was reversed when the defendant was Black (*M* = 3.52 for attractive victims, *M* = 5.01 for unattractive victims)
28. Piatak (2015)	Sexual offence	Blameworthiness: (8-point scale) indicating the degree to which the mock-juror agreed with each statement.	Blame: the effect of victim race nonsignificant. Effect of mock juror race significant. The mean blameworthiness score for White mock jurors (*M* =15.33, *SD* = 4.53) was significantly lower than the mean score for Black (*M* =17.95, *SD* = 3.89) and other race mock jurors (*M* =19.04, *SD* = 4.09).

Findings.

Verdict.

Defendant race. The findings on the possibility of racial bias against defendants in verdicts were mixed. Eighteen out of 25 studies found no significant effect of defendant race on mock jurors' decisions of guilt ([3,4,5, 9, 10, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 25, 26, 27]). Seven studies found a significant effect of defendant race on the verdict ([1,2,6 8,11,19,24]). Specifically, five studies demonstrated that mock jurors were more likely to find White defendants guilty than Black defendants ([1,2,6,8,19]); whilst in two studies Black defendants were more likely to be convicted than White defendants ([24,26a]).

Interactions between defendant race and verdict were found. One study found Black mock jurors were significantly less likely to convict Black defendants identified by a White eyewitness of second-degree murder, suggesting ingroup racial bias ([14]). One study found defendant race and intellectual disability significantly interacted to impact on verdict, with mock jurors more likely to find White non-disabled defendants guilty of first-degree murder compared to Black non-disabled defendants ([6]). Interestingly, one study found a significant interaction between the race and mental disorder, such that Black defendants with depression were six times less likely to be found guilty than Black defendants with schizophrenia ([5]).

Victim race. No main effect of victim race on verdict was found. However, one study found that defendant and victim race interacted, with mock jurors being least likely to convict Black defendants accused of impaired driving whose victims were White, compared to other victim-defendant race combinations ([12]). In contrast, one study ([9]) did not find a significant interaction between defendant and victim race. However, the interaction of defendant race, victim race and victim age were significant in this study. Mock jurors attributed less guilt to the Black defendant than the White defendant when the victim was 17 years old and Black.

Punishment/ sentencing.

Defendant race. Seven out of eleven studies reported a non-significant main effect of defendant race on mock jurors' sentence recommendations ([1,3,9,13,15,16,27]). However, four studies reported a significant main effect of defendant race on mock juror decisions on punishment, with Black defendants receiving more lenient sentences than White defendants ([7,11,12,20]).

In addition, two studies found significant interactions between defendant race and race saliency ([15]); and socioeconomic status (SES) and level of physical attractiveness ([16]). McManus et al. (2018) reported that mock jurors were more likely to recommend severer sentences to Black defendants accused of dangerous operation of a motor vehicle and impaired driving causing bodily harm, in high race saliency versus low race-saliency conditions. However, sentencing decisions for White defendants did not differ by race saliency condition. Elsewhere, Armenta (2017) found that mock jurors recommended lengthier sentences to Black defendants accused of murder (filicide: mother against child) with low SES and less physical attractiveness compared to defendants with other combinations of race, SES, and physical attractiveness.

Victim race. No main effect of victim race on sentence was found. Two studies examined the influence of victim race on mock jurors' sentencing decisions ([7,9]). One study found mock jurors were

significantly more likely to recommend death when the victim was Black than White ([7]) whereas one study did not find a significant effect of victim race on mock jurors' sentencing recommendations ([9]).

Blame.

Defendant race. The findings on the influence of defendant race on mock jurors' attribution of blame were mixed. Five out of nine studies found that mock jurors were significantly less likely to blame Black than White defendants for an offence ([2,4,8,9,15]). One study found mock jurors were significantly more likely to blame Black than White defendants for the crime of murder (filicide) ([16]); and three studies found no significant main effect for defendant race on the attribution of blame for the crimes of dangerous operation of a motor-vehicle (DOM) and fraud, first degree murder and robbery, and robbery and theft of a motor vehicle ([22,23,27]).

Note, interactions between defendant race and blame were found. Maeder et al. (2015) and Maeder et al. (2016) found mock jurors' decisions on blame significantly interacted with the type of crime committed by the defendant and the victim's level of physical attractiveness. Specifically, mock jurors considered White defendants more culpable for certain crimes (e.g., fraud) and Black defendants for others (e.g., auto-theft) ([22]); and considered White defendants less culpable when the victim was attractive (compared to unattractive), whilst Black defendants more culpable when the victim was unattractive ([26]). Interestingly also, Yamamoto and Maeder (2017) found evidence suggesting ingroup racial bias by Black mock jurors, who tended to perceive Black defendants accused of looting as being in less control of their actions than did White jurors.

Victim race. The findings on the influence of victim race on mock jurors' attribution of blame were mixed. Two out of three studies found that the victim's race did not significantly affect the culpability decisions of mock jurors for a sexual offence crime ([9,28]). However, the third study found a significant interaction between defendant and victim race ([2]), with mock jurors considering victims

more blameworthy for a child sex trafficking offence when the defendant was Black than White; and considering White defendants as more blameworthy than Black defendants. Defendant race rather than victim race appeared to be influential in mock-jurors' perceptions of blame.

Quality assessment

Studies were critically appraised using the Center for Evidence Based Management (CEBMa) appraisal tool for cross-sectional study (study) (CEBMa, July 2014). The findings are summarised in appendix D.

All studies had clear aims and used methods appropriate for their research question. All studies clearly described the participant selection method, however, most of the included studies were subjected to selection bias. Despite that, half of the included studies (*n* = 14) had a relatively representative sample but only four studies reported the sample size estimation procedure ([1,5,9,25]). None of the included studies reported on the response rate and only three studies presented information about the validity or reliability of their measurement tools ([5,22,28]). All studies assessed the statistical significance as part of the data analysis but only seven studies ([3,6,9,15,19,20,24]) reported 95% confidence interval for the observed effects. None of the studies adjusted for confounders other than the parameters included in the prespecified factorial design. Since none of the studies were conducted outside of the US and Canada, or specifying individual professional groups (e.g., criminal justice professionals) it is also difficult to generalise the findings on the racial bias among the jurors to other countries and to the decisions made by people of varying professions. Although collectively most studies met most of the conditions in the CEBMa tool, there was clear evidence of bias in the selection process, lack of reporting of response, and inadequate information about the validity or reliability of the assessment methods. Moreover, the representativeness and generalisability of most studies were questionable.

Discussion

This systematic review explored evidence on the differential treatment of Black versus White defendants and victims by mock jurors on decisions of guilt, blame, and punishment.

Overall, mock-juror decisions on verdict were not significantly influenced by the defendant or victim's race. However, mock jurors' verdicts were influenced by the type of crime committed, and non-racial characteristics of the defendant or victim, including mental health disorder, intellectual disability, and age. This relationship was further modified by extra-legal factors, including race saliency, physical attractiveness, and SES. The effect of defendant race on the attribution of blame by the mock-jurors was more complex, with the type of committed crime being a key contributor.

Verdict

Defendant race. The present systematic review did not find clear evidence of discrimination against Black defendants amongst mock jurors on measures of guilt. Most included studies did not find a significant main effect of defendant race on verdicts, which is consistent with the findings of some previous meta-analyses (Devine & Caughlin 2014; Mazella & Feingold 1994). However, some studies included in the current review found race of defendant significantly influenced verdict, which some studies finding White defendants were significantly more likely to be convicted than Black defendants ([1,2,6,8,19]), whilst other studies found the opposite ([24,26]).

Of relevance to the current systematic review is the criticism that the Mazella and Feingold (1994) did not consider the possibility that the race of the participant (mock-juror) might influence verdict. The current review also did not explore the influence of mock juror race on verdict. Previous meta-analyses

that have considered the race of mock juror on verdict have found mock jurors more likely to give guilty verdicts to defendants who were of a different racial group to their own (Mitchell et al., 2005; Devine & Caughlin, 2015). Specifically, the meta-analysis by Mitchell et al. (2005) included 34 studies and found racial bias was more evident in Black than White mock jurors; whilst Devine and Caughlin (2015) meta-analysis included 32 studies and found that Black and not White mock jurors showed ingroup favouritism in their verdict decisions. The implication for the current review is that if the review question included the influence of participant (mock-juror) race on verdict decisions, then perhaps similar ingroup biases to that of previous meta-analyses could have been found. Consequently, an inclusion criterion for future systematic reviews is that the race of the participant (mock-juror) on decisions is assessed.

Victim race. None of the studies in this review explored the main effect of victim race on verdicts. There was mixed evidence for an interaction between victim and defendant race on mock juror verdicts. While Call et al., (2019) did not find a significant interaction between victim and defendant race, Maeder and Yamamoto (2019b) reported a significant interaction. In their study, Black defendants with White victims were least likely to be convicted for impaired driving, compared to all other victim-defendant combinations. The finding contradicts some earlier research where mock jurors were more lenient toward White defendants with Black victims for crimes of murder and sexual assault (Baldus et al., 1990; Bohm, 1994; Keil & Vito, 1995). Also, there is evidence of a significant interaction between mock juror, defendant, and victim race in relation to the offence of child sexual assault (Foley & Chamblin, 1982). The researchers showed that White mock jurors assigned higher guilt ratings to a Black versus White defendants irrespective of the victim race. In contrast, Black jurors assigned higher guilt ratings to the defendants with a same race victim, irrespective of the defendant-victim race combination (Foley & Chamblin, 1982). It is worth noting that when the latter study

statistically controlled for SES, the effect of race on verdict was eliminated. However, none of the three studies included in the present systematic review ([14,17,20]) found a significant interaction effect of mock juror, defendant, and victim race on verdicts.

Interactions between defendant and victim race. This review found that mock jurors' verdicts were influenced by the type of crime committed, mental disorder, intellectual disability, victim age, and race of eyewitness ([5,9,6,14]). Two studies in this review found a significant interaction between crime type and defendant race ([3,22]). Further analysis of the significant results in this review has found that the race of defendant seemed to interact with the type of crime to influence verdict. Note, in five studies, mock jurors were more likely to convict White than Black defendants of assault, child sex trafficking, first degree murder, and residential burglary ([1,2,6,8,19]). In comparison, two studies ([24,26]) reported that Black defendants were more likely to be convicted of triple murder and robbery and theft of a motor vehicle. This finding is supportive of early studies on race-crime congruency, which found that defendants accused of committing crimes considered congruent to their racial group tended to be treated more harshly than defendants accused of crimes rendered racially incongruent (Gordon et al., 1998; Jones & Kaplan, 2003; Skorinko & Spellman, 2013).

The notion of race-crime congruency has received some meta-analytic support (Devine & Caughlin, 2014). It is thought that racially biased decisions occur when there is a match between the defendant's race and the crime considered stereotypical for the defendant's racial group (cf., Gordon et al., 1998). Therefore, race-crime incongruency might be a possible explanation for nonsignificant results in some studies. Furthermore, it is worth noting that the meta-analysis of Mazella and Feingold (1994) which included 29 studies, found that the effect of race on guilt was not significantly moderated by the type of crime. However, an explanation for the discrepancy between the current finding and that of Mazella

and Feingold, might be due to the present systematic review including a broader range of crime types to examine the influence of race on verdict. Additionally, the difference in findings might reflect a change of attitudes and/or race-crime stereotypes over time.

Punishment/ sentencing.

Defendant race. There was some evidence of less severe punishment decisions against Black versus White defendants ([7,11,12,20]). Four studies found a significant main effect of defendant race on mock juror sentencing decisions regarding murder ([7]), robbery ([11]), impaired driving; ([12]), and looting ([20]). The finding is contrary to a previous meta-analysis (Sweeney & Haney, 1992) which demonstrated that mock jurors treated Black defendants more harshly in their sentencing decisions than White defendants. However, Sweeney and Haney (1992) reviewed studies that only included White mock jurors. Notably, two later meta-analyses by Devine and Caughlin (2014) and Mitchell et al., (2005) which included views of both Black and White mock jurors, found ingroup favourable bias towards defendants of the same race. Consequently, it seems important that future studies on racial bias assess the influence of mock jurors' race on sentencing decisions.

Victim race. None of the studies in the review examined the impact of victim race on sentencing decisions by mock jurors. However, two studies explored the interaction between victim and defendant race on punishment ([7,9]). Bell (2019) reported that mock jurors tended to choose the death penalty over Life without Parole (LWOP) for the offence of murder when the victim was Black versus White. This conflicts with previous simulated and archival data that showed defendants accused of murdering Black victims received the death sentence less frequently than defendants with White victims (Baldus et al., 1998; Bohm, 1994; Keil & Vito, 1995). In contrast, Call (2019) did not find a significant interaction between victim race and mock jurors' sentencing decisions for a sexual offence.

This nonsignificant interaction is supported by a previous meta-analysis (Mazella & Feingold, 1994) which demonstrated that mock jurors recommended severer punishments when the victim was White rather than Black. Importantly, older studies suggested that, in rape cases, White mock jurors were more punitive when the defendant was Black and the victim was White compared to other combinations of defendant and victim race (Feild, 1979; Foley & Chamblin, 1982; Klein & Creech, 1982; Ugwuegbu, 1979). However, these studies are dated, and racial attitudes might have changed over time, with a possible decline or reversal of racial bias.

Punishment and other interactions between defendant and victim race. In addition to the above interactions, the present review found race saliency, physical attractiveness and SES as other factors that significantly interacted with defendant and victim race on mock jurors' sentencing decisions ([15,17]).

McManus et al. (2018) showed that mock jurors would recommend a severer sentence for the DOM and impaired driving causing actual bodily harm to Black defendants whose race was made salient. This contrasts with several US based studies, which have suggested less racial bias in White mock jurors when defendant race was made salient (Sommers & Ellsworth, 2001). A possible reason for this is socially desirable responding, where participants try to appear unbiased (Axt et al., 2016). Perhaps then, when it comes to studies in this review that were conducted in the US, the null findings might be influenced by race being made salient.

Interestingly, in the current systematic review, one study found a significant interaction between defendant race, SES, and physical attractiveness ([16]). Armenta (2017) found that Black defendants with low SES and rated as unattractive, were recommended lengthier prison sentences than their White counterparts. Such findings highlight the complexity of the issue of race and mock-juror decision

making, particularly that other extra-legal factors aside from race might play a role in mock juror decisions.

Blame

Defendant race. There were more mixed findings for attribution of blame by mock jurors in the included studies. Four studies found mock jurors were more likely to blame White than Black defendants for child sex trafficking ([2]), assault ([4]), residential burglary ([8]) and DOM and impaired driving causing bodily harm ([15]). One study found that Black defendants were considered more blameworthy for the offence of filicide ([16]). Four studies found no significant main effect of defendant race on mock jurors' attribution of blame for sexual offending ([9]), dangerous operation of a motor vehicle while intoxicated/ fraud ([12]), first degree murder ([23]) and armed robbery ([26]).

Previous US based studies have found that in reference to rape, Black men accused of rape were considered more culpable when their victims were White compared to when the victim was presented as Black (George & Martinez, 2002; Ugwuegbu, 1979; Varelas & Foley, 1998). However, these studies are dated, and attitudes and behaviour might have changed over time. Unfortunately, no previous systematic reviews or meta-analysis have been done in this area to compare these findings with. This is an area for future research.

Victim race. One study showed mock jurors attributed more culpability to victims of Black versus White defendants for the crime of child sex trafficking ([2]); whilst two studies reported that race did not influence mock juror decisions on blame for a sexual offense ([9,28]).

Interestingly, there has been research indicating that mock jurors (i.e., a student sample) were less likely to consider a rape scenario as a crime when the victim was described as Black compared to when they were described as White (Foley et al., 1995).

No previous systematic reviews or meta-analysis have been done in this area. However, in terms of sexual assault at least, there is research indicating that White victims of a Black defendant were seen less blameworthy for the offense, compared to Black victims of a White defendant (Bottom et al., 2004; George & Martinez, 2002), whilst other studies have found the opposite effect (cf. Allison & Wrightsman, 1993).

Nevertheless, as there are few studies in this review that deal with this topic, it is difficult to draw conclusions.

Blame and other interactions between defendant and victim race. As with the finding on verdict, this review found evidence for an interaction between mock jurors' culpability judgements and crime type ([22]). Specifically, White defendants were considered more blameworthy for the crime of fraud than dangerous operation of a motor vehicle. Contrastingly, Black defendants had the highest blame ratings in the auto theft trial than the fraud and dangerous operation of a motor vehicle trials. The results provide modest evidence of a race-crime congruency effect.

One study found a significant interaction between victim physical attractiveness and blame, for a sexual offence ([27]). When the defendant was White, attractive victims received higher responsibility ratings than did unattractive victims. The opposite effect was found for when the defendant was Black.

No previous systematic reviews or meta-analysis have been done in this area to compare the results of this systematic review with. The lack of studies within this area makes it difficult to make generalisations. This is an area for future research.

Furthermore, it is unclear how blame attribution is associated with the verdict and sentencing decisions of mock jurors. More research needs to be done exploring the relationship between the degree of blame attributed to the victim/defendant, and mock jurors' guilt and sentencing decisions.

Some practical implications of the findings. The mixed evidence from this systematic review on race and juror decision making, suggest that it is inconclusive as to whether mock jurors are racially biased in their decisions on guilt, sentencing and blame. Consequently, alternative explanations for the racial disproportionality in the criminal justice system to that of racial biased officials have been given. Regarding sentencing, it is important to note that in the UK sentence type and length is affected by whether a defendant pleads guilty. Judges can reduce the sentences of offenders by up to one third if they plead guilty. Note, in the UK the Lammy (2017) review found that Black defendants more often pleaded guilty compared to White defendants. Consequently, if Black defendants plead not guilty but are found guilty, the option of their sentence being is reduced is ruled out. Therefore, racial discrepancies in the sentencing of Black versus White defendants might be the result of Black defendants' plea decisions, rather than a racially biased criminal justice system. However, in terms of guilt decisions, it is possible that Black defendants are simply more likely to be found guilty than similarly situated White defendants, even when they are innocent.

Intersectionality has been used to explain racial disproportionality in the CJS. Intersectionality asserts that people consist of multiple identities (based factors including race, gender, class, socioeconomic status) that intersect and impact on their experience of inequalities and injustices (Crenshaw, 1989).

No single factor is thought to contribute to an injustice. Perhaps then, Black people are more likely to have additional identifies that disadvantage them in the criminal justice system (e.g., being of low socioeconomic status), which collectively account for their disproportionality.

Research limitations

The limitations of this review concern the recruitment procedures, participants, methodology and statistical procedures employed in some of the included studies.

Recruitment. Although most included studies had participants that met the criteria for jury eligibility, studies either recruited participants from universities or via online platforms (e.g., MTurk), which might limit the representativeness and generalisability of the findings (Berinsky et al., 2012; Huff & Tingley, 2015; Nunan et al., 2018; Paolacci et al., 2014). Furthermore, some studies recruited participants from only one university or course, which might introduce geographical bias, which some research has found significant when studying racial bias (Borg, 1997). Similarly, although using online platforms might improve the representativeness of the study sample compared to using students, some platforms (e.g., MTurk) have online discussion boards, which could be used to share information about studies thereby reducing the naivety of participants, biasing their responses and ultimately the overall quality of the data (Aguinis et al., 2021; Chmielewski & Kucker, 2020).

It is possible that people who drop out of research are significantly different in terms of attitudes and behaviour from those who remain. Only one study in this review reported on attrition. Without attrition figures it is difficult to ascertain whether studies were affected by self-selection/ attrition bias.

Population. All included studies used either US or Canadian citizens as participants, with most being students, White and female. Some studies have found that students tend to have more liberal attitudes than the general population, which might lead them to significantly different legal decisions from the general juror population (Eagan et al., 2015). Similarly, some studies including meta-analysis have found females to be more punitive for certain offences (particularly sexual) than males (Devine & Caughlin 2014; Payne & Chappell, 2008). Perhaps then, a more diverse sample recruited from a wider geographical region and continents, a non-student population, and using a variety of recruitment methods, might have produced different results (cf. Armenta, 2017).

Methodological. A major methodological limitation is the use of the mock-jury paradigm itself. First, although the mock-juror paradigm has many advantages (e.g., it is easy to replicate), its use of transcripts, audio recording and video clips is very different from a real trial setting, reducing its ecological validity (Bornstein et al., 2017). Second, the variation in the range, type and quality of the stimulus materials used across studies dilutes the homogeneity of them, and such differences might account for variability in findings. Third, individual mock-juror decisions might be different from jury (group) decisions, or juror decisions following jury (group) deliberation (Haegerich et al. 2013; Kerr et al., 1996). Fourth, the fact that participants were aware their decision had no real-life consequence might have affected their decisions and be dissimilar to decisions made in actual jury settings (Bornstein et al., 2017). Collectively, the latter factors ultimately impact on the overall generalisability of these studies.

Other factors that must be considered in interpreting the findings relate to the type of outcome measures used across the studies. Regarding differences in results because of using either

dichotomous versus continuous scales, the findings are mixed (Devine & Caughlin, 2014; Mitchell et al. 2005). Whilst Devine and Caughlin's (2014) meta-analysis found continuous verdicts produced a slighter weaker effect than dichotomous verdicts, Mitchell et al. (2005) found the opposite. Mitchell et al. (2005) suggests that if procedures in mock-juror studies were made more ecologically valid, then this might reduce observed racial bias (cf. Pfeifer & Ogloff, 1991).

Finally, most included studies did not report that the outcome measures used to assess guilt, blame or punishment were reliable or had been validated. This potentially limits the findings.

Statistical procedures. Statistical issues concerning the categorisation of participants might also be a limitation. Some studies grouped certain racial groups together for statistical analysis, which might limit the ability to detect diversity of opinions within a sample. Some studies did not control for juror race, which previous meta-analyses have found influences mock-juror decisions – specifically in terms of ingroup/ outgroup racial bias (Devine & Caughlin 2014; Mitchell et al. 2005). Therefore, some of the observed differences might reflect outgroup/ingroup bias rather than specific views held about people. The meta-analysis by Devine and Caughlin (2014) found that Black and not White mock-jurors showed favourable ingroup bias toward defendants who were of their racial group. The latter is important as perhaps some of the null studies reflected the higher use of White mock-jurors who are less likely to show racial bias than Black mock-jurors. If studies controlled for race of mock-juror, then racial bias by Black mock-jurors would have been found.

Finally, as indicated in the results section, only four studies explicitly stated that their sample size was based on pre-study considerations. Perhaps the decisions of jurors were influenced by race, but due to studies being underpowered this was not statistically detected.

Quality assessment tools. The current systematic review is not without limitations. The use of CEBMa critical appraisal tool is a possible limitation. Nevertheless, the tool does cover most of the criteria found in other critical appraisal tools and assesses more domains than the more familiarly used CASP (Critical Appraisal Skills Program) tool. Therefore, the CEBMa can be considered a compromise between longer critical appraisal tools such as the Appraisal Tool for Cross-section Studies (AXIS), and shorter appraisal tools such as the CASP.

The systematic review's date inclusion criterion. As stated earlier an inclusion criterion for this systematic review was that studies needed to be published between January 2015 and November 2021. Note, the last meta-analysis was conducted by Devine and Caughlin in 2014. Consequently, it is possible that the current systematic review missed studies that might have been conducted between the end date of the previous meta-analysis and the start date of the current systematic review. Furthermore, there would be a time lapse between the previous meta-analysis being conducted and its publication date. Consequently, to increase the possibility of capturing any missed studies, it would have been better if the start date for studies included in the current review had been before the publication date of the previous meta-analysis.

Recommendations. Based on the findings from this review, its comparison to other research in the field, and the review limitations, it is recommended that:

- mock juror studies are conducted outside of the US and Canada, as race of defendant and victim might affect mock jurors from other countries differently from those in the US and Canada.
- study participants should be recruited outside of online platforms and universities, which might increase the representativeness of the sample to the target population.
- research samples should have a more equal gender and race balance – and include in their analysis responses by gender and race, so to assess whether these variables influence mock juror decisions.
- studies should cover the influence of victim blame on mock jurors' verdicts and sentencing decisions.
- future systematic reviews and/or meta-analyses should explore the influence of race salience, crime type, physical attractiveness, SES, and mock juror race.
- future reviews might focus on mock jury decision making, as decisions made in groups might vary from individual decisions.
- future systematic review must ensure the inclusion dates for studies included in the review adequately covers the period between the last systematic review/ meta-analysis date and the new review, to avoid the possibility of missing studies that should be included in the review.

Conclusion. This systematic review addressed whether mock jurors treated Black defendants and victims differently from White defendants and victims on decisions of guilt, blame and punishment. The findings were mixed.

Most included studies found no significant main effect of defendant race on mock jurors' decisions of guilt, and none found a main effect of victim race on guilt. Such findings might reflect that most mock jurors are not racially biased in their decisions on guilt, and that victim race does not significantly influence guilt decisions.

Most studies (4/7) found a significant main effect for defendant race on mock juror sentencing severity, with Black defendants tending to be recommended more lenient sentences than White defendants; no main effect of victim race on sentencing decisions were found. There was mixed evidence for the influence of race of defendant and victim on mock jurors' decisions on blame. Such findings seem to reflect that where there are biased sentencing decisions, Black defendants seem to be treated more leniently than White defendants, and that victim race does not significantly influence mock juror decisions. These findings contradict figures on the disproportionate number of Black compared to White people entering the criminal justice system. The finding from this systematic review suggests racial bias is less evident at juror level, suggesting that the racial disproportionality figures in the CJS do not reflect decisions made by jurors. Perhaps racial bias occurs elsewhere within – or at other stages of the criminal justice system. Perhaps more research needs to look at other areas of the criminal justice system (e.g., decisions around arrests and prosecution) to explore whether they are racially biased.

Finally, it is worth noting that the review found that some studies highlighted there are several extra-legal factors including type of crime that can influence mock jurors' decisions on verdict, punishment, and blame. Therefore, although the findings in the studies from this review might reflect genuine attitudes about defendant and victim race, they might have also been confounded by various extra-legal factors. The latter alongside factors such as the way some samples were obtained, the representative of the sample, the same size, the response rate achieved, and the uncertainty around the reliability and validity of some of the measurement tools used, question the quality of some of the studies included in this review. The findings from this review should be interpreted with caution, and future research needs to consider such factors when conducting research.

Chapter 3: Measuring Socially Desirable Responding: A Critique of the BIDR-16

Introduction

The previous chapter highlighted that risk of bias is a concern for researchers. Socially desirable responding (SDR) is a form of response bias where the respondent is inclined to provide positive self-descriptions (Johnson & Fendrich 2005). A particular concern for researchers using self-report measures, such as those used in simulated jury/juror research (and the current thesis), is that respondents may provide untruthful answers which distort and invalidate their results (Leite & Beretvas, 2005; Paunonen & LeBel, 2012). Consequently, researchers have found ways to detect and control for SDR, including developing standalone SDR psychometric tests (Paulhus & Vazire, 2007).

This chapter critiques the standalone SDR measure known as the Balanced Inventory of Desirable Responding-16 (BIDR-16) developed by Hart et al., (2015) (see appendix E for a copy of the BIDR-16). The BIDR-16 is a refined version of the Balanced Inventory of Desirable Responding (BIDR), developed by Paulhus (1984, 1991, 1998). The BDR-16 is used in the empirical studies discussed in chapters four and five of this thesis, to assess for socially desirable responding amongst criminal justice workers. Consequently, the BIDR-16 was critiqued here to ensure it has good psychometric properties of validity, reliability, standardisation, and practicality. Furthermore, the lack of critical evaluation of the BDR-16 supported the need for this critique, as did the increased use of shortened SDR scales to combat SDR. As the BIDR-16 was developed using the full scale BIDR measure, the critique will initially provide a brief overview and evaluation of the original BIDR.

The Balanced Inventory of Desirable Responding (BIDR)

The original BIDR (Paulhus, 1984) was developed when controversy around the dimensionality of SDR emerged. Several measures considered SDR a unitary construct, including the Marlowe-Crowne Social Desirability Scale (MCSDS) (Crowne & Marlow, 1960) and the Eysenck Lie Scale (Eysenck & Eysenck

1964). However, low correlations between different SDR measures (as well as several factor analytic studies) suggested that SDR could be a two-factor or multi-factor construct (Barger, 2002; Milliam & Jacobson, 1978).

Research by Paulhus (1984) supported the partition of SDR into two different but related clusters. Cluster one links to the MMPI's Alpha factor (Block, 1965), and represents an unconscious positive bias in self-reporting that appears linked to specific personality constructs including Openness and Extraversion (Scheier & Carver, 1985; Taylor & Brown, 1988). Cluster two factors are linked to the MMPI's Gamma factor including Conscientiousness and Agreeableness (Wiggins, 1964). The Gamma factor is seen as a more conscious bias in self-reporting aimed at and affected by an audience. Emotional stability is mostly related to the alpha factor but is also strongly correlated with gamma (Paulhus & John, 1998).

Considering factor analytic evidence on the dimensionality of SDR, Paulhus developed the BIDR measure. The BIDR is a 40-item, two-factor measure of SDR designed to capture the alpha and gamma factors. It comprises evenly balanced attribution and denial items, split across two subscales (i.e., 20 items per scale): self-deceptive enhancement (SDE), which captures the Alpha factor; and impression management (IM), which captures the Gamma factor. SDE refers to the respondent's denial of undesirable self-descriptions and is thought to be primarily unconscious. IM refers to the respondent's inclination to consciously create favourable impressions of themselves on others and can validity-check respondents' likelihood of faking responses (Paulhus, 1994).

The full version of the BIDR is the BIDR-7 (Paulhus, 1998), also named the Paulhus Deception Scale. Differences between version 6 and 7 is that the BIDR-6 used a 7-point scoring scale whilst the BIDR-7 used a 5-point scoring scale; and in version 6 the SDE formed part of a self-deceptive positivity subscale that comprised of SDE alongside a self-deceptive denial (SDD) subscale (i.e., a further 20 scale items).

The SDE subscale closely resembled the original Self-Deception Questionnaire (SDQ) developed by Sackeim and Gur (1978) and based on the psychoanalytic assumption that many people cannot handle their normal sexual and violent fantasies and depress their less attractive side. However, Paulhus and colleagues found it impossible to separate denial from reality and abandoned the SDD subscale (Paulhus & John, 1998; Paulhus & Reid, 1991). Consequently, the BIDR-7 reverted to having just the IM and SDE subscales.

In terms of completion, the BIDR-7 requires respondents to use a 5-point Likert type scale to rate (from 1 (not true) to 5 (very true)) the extent that each item/statement applies to them. The scale can be scored continuously or dichotomously. Scores range from 0 to 40. Each subscale is scored separately. The overall scale score is the sum of the two subscales.

The reliability of the BIDR. Paulhus (1998) demonstrated acceptable reliability for the SDE subscale, reporting Cronbach's alpha coefficients ranging from .70 (college students) to .75 (general population). On the IM subscale, good Cronbach's alpha coefficients were reported, with scores ranging from .81 (college students) to .84 (general population). Similarly, the reported reliability for the total scale was good, with Cronbach's alpha coefficients scores ranging from .83 (college students) to .86 (prison entrants).

Over a one-week interval, acceptable test-retest reliability[1] coefficients for dichotomous and polytomous scoring were found for the SDE subscale (.78 and .83), and good reliability coefficients were found for the IM subscale (.83 and .86) (Vispoel & Tao, 2013). Over a five-week interval, questionable test re-test reliability coefficients were found for the SDE (.69) and IM (.65) scale scores

[1] test-retest reliability – is consistency over time. For example, if the same instrument was used with the same study population at a different time, would the results be like the original result.

(Paulhus ,1991). Over a longer five-month period, a test-retest reliability coefficient of .74 was reported for the IM scale score (Crutzen & Goritz, 2010). However, over a two-year and three-year interval, stability coefficients were found to be questionable to unacceptable for both SDE (.71 and .44 respectively) and IM subscales (.68 and .68 respectively) (Lönnqvist, et al., 2007).

A meta-analysis by Li and Bagger (2007) that used 110 studies, found the BIDR had adequate reliability. The mean score reliability coefficient of the IM scale was .74 (*SD* = .09, *n* = 107). The mean reliability coefficients were .68 (*SD* = .09, *n* = 90) for SDE scores and .80 (*SD* = .04, *n* = 18) for overall BIDR scores.

The validity of the BIDR. In terms of convergent validity[2], Paulhus (1998) reported positive correlations between the SDE subscale and the Edward Social Desirability measure (Edwards, 1957), MMPI K (Meehl & Hathaway, 1946) and Personality Research Form (Jackson, 1984) scales. Barchard (2002) reported a small positive correlation (.24) between the SDE subscale and the Positive Expressivity Scale (PES). A positive intercorrelation was also reported between the IM and the Social Desirability (Wiggins, 1959), Eysenck Personality Inventory (Eysenck & Eysenck, 1964) and MCSDS scales. An earlier report by Paulhus (1988), found that the total SDR scale correlated with the MCSDS and Multidimensional Social Desirability Inventory (Jacobson et al., 1977) (.71 and .80 respectively). These findings support for the measure's convergent validity.

In relation to discriminant validity[3], moderate correlations between the SDE and IM have been reported (from .05 to .40) (Paulhus, 1991, 1998). The finding that SDE and IM loaded on different

[2] convergent validity – exists if the correlation coefficient for two instruments/scales measuring a similar construct s(e.g., depression) is > .50, or preferably > .75 (Carlson & Herdman, 2012).

[3] Discriminant validity refers to how much two instruments/scales that are supposed to measure different constructs are uncorrelated. If the correlation between these measures is weak and non-significant then discriminant validity is likely to exist. (Carlson & Herdman, 2012).

factors, support the discriminant validity of the scales and has been used to support the two-factor model of SDR (Paulhus, 1984, 1998). Specifically, SDE has positively correlated with narcissism, overconfidence, perceived control, optimism, self-efficacy, individualism, and the Big Five personality traits of Contentiousness and Extraversion (Paulhus, 1991, 1998a, 1998b); and negatively correlated with neuroticism. IM scores strongly positively correlated with the Big Five traits of Agreeableness, Conscientiousness and Openness; and with dissimulation, including responsiveness to situational demands (Holden et al., 2003; Lalwani, et al., 2006; Paulhus, 1994). As a respondent's anonymity decreased, their scores on IM (but not on SDE) increased (Paulhus, 1994). Similarly, adjustment positively correlated with SDE and not IM, whilst SD (but not IM) predicted overclaiming, overconfidence and hindsight (Paulhus, 1994). The construct validity[4] of the BIDR at item level has also been supported by factor analytic research (Holden, Starzyk, McLeod, & Edwards, 2000).

Limitations of the BIDR 40. The BIDR has not been without criticism. Specifically, the BIDR-7 manual fails to report the factor loading (i.e., construct validity) associated with the two-factor model and model index values. Furthermore, although Paulhus has allowed the BIDR to be scored using polytomous (i.e., items that have more than two possible scores e.g., Likert scale) or dichotomous scoring methods, there is conflicting results on which method is preferable. Dichotomous scoring leads to a loss of information in terms of individual differences, as only extreme responses are scored (MacCallum et al., 2002). Polytomous scoring has improved reliability and convergent validity (Stöber et al., 2002), and item response theory analysis supports the use of polytomous scoring (Vispoel & Kim,

[4] construct validity – is the degree that a measurement (e.g., BIDR-16) measures the construct that it is designed to measure. For example, does the BIDR-16 truly measure socially desirable responding or does it measure something else? A way to verify construct validity is to compare the test to other tests that claim to measure to same construct to see how much they correlate.

2014). However, Leite and Beretvas (2005) did not support a two-factor model fit derived from 7-point rated polychoric scores (CFI = .614, TLI = .593), which led them to support dichotomous scoring. Leite and Beretvas's proposal has received support (Stöber et al., 2002) and led them to conclude that the BIDR should be used with caution.

Another issue with the BIDR has been its length. If the BIDR is administered alongside a series of other measures, then it may increase the possibility of transient errors (e.g., fatigue) impacting on the completion of questionnaires/psychometric instruments (Schmidt et al., 2003). Consequently, some psychometricians have worked on developing short scales that improve the practicality of the original measure.

Despite the controversy around the dimensionality of SDR and the scoring and length of the BIDR, there may still be value in considering and assessing SDR in terms of SDE and IM and the two-factor polytomous BIDR model (Asgiersdottir, et al., 2016; Stober, 2001).

The Balanced Inventory of Desirable Responding-16 (BIDR-16)

Hart et al. (2015) developed the BIDR-16 to improve the practicality of the original BIDR, whilst retaining its two-factor structure, reliability, and validity via four sequential studies. Study one evaluated the construct validity of the original BIDR (after removing four extremely variable items) and refined it into a shorter scale whilst maintaining the fitness and dimensionality of the original scale. Study two explored the independent construct validity of the BIDR-16 in a different sample. Study three examined the retest reliability of the new BIDR-16 scale. Study four tested the external validity[5] of the BIDR-16 scale against other SDR measures. The findings from these studies are discussed below.

[5] external validity – Is the extent that results from a study can be generalised to wider contexts (e.g., people events, situations, groups).

Study one

Assessing the validity of the original BIDR. The researchers used the 7-point polytomous scoring of the BIDR-40 instead of dichotomous scoring recommended by Stöber et al., (2002) and Kam (2013). In terms of construct validity, CFA[6] of the original two-factor BIDR was undertaken on adults (1,479 females and 465 males) predominately from the UK (*n* = 854) and USA (*n* = 815). The CFA found undesirable or marginal overall fitness for this model (GFI = .89, CFI = .84). However, when assessed by indices that were robust to sample size and factor loading, more acceptable model fitness (RMSEA = .058) was observed (see Hu & Bentler, 1999). Table four provides an overview of fit indices for different factor solutions within CFA.

[6] Confirmatory factor analysis (CFA) is a statistical procedure employed to substantiate the factor structure of a set of observed variables. The technique lets the research test whether a relationship exists between observed variables and their underlying latent constructs

Table 4

Goodness-of-fit Indicators of models for a 36-item, 20-item, and 16-item two-factor BIDR model

Model	*χ2*	GFI	CFI	RMSEA	SRMR
36-item 2-factor model	3,627.04	.89	.84	.06	.06
20-item 2-factor model	1,417.00	.92	.88	.07	.06
16-item 2-factor model	913.92	.93	.88	.07	.05

Note. The values are taken from the Hart et al., (2015) paper. Note also the authors did not report *df, χ2 /df* and *χ2 diff* for chi-square.

Refining the original BIDR scale. The researchers reduced the size of the original scale to a model with eight items loading on each factor. The refined model provided acceptable fit (GFI = .93, CFI = .88, RMSEA = .07). The goodness of fit for the refined 16-item two-factor model (BIDR-16) was tested across four subsamples (i.e., UK and US nations; and by male and female genders). The refined model fit reasonably well in all subsamples. See table 5.

Table 5

Goodness-of-fit indicators for the BIDR-16 by nation and gender.

Sample	χ^2	GFI	CFI	RMSEA
UK (814)	471.50	.93	.84	.07
US (778)	429.30	.93	.91	.07
Women (1,405)	672.17	.92	.90	.07
Men (440)	331.95	.92	.83	.07

Note. The values are taken from the Hart et al., (2015) paper. The authors did not report *df,* χ^2 */df* and χ^2 *diff* for chi-square.

Evaluating the external validity of the BIDR-16 against the original BIDR scale. As a secondary aim of study one, correlational tests were performed to evaluate the external validity of BIDR-16 subscales against the corresponding subscales of the original BIDR and self-enhancement (self-esteem and narcissism). See table six.

Table 6

Correlations between the BIDR, BIDR-16, and existing scales in study 1.

Scale	SDE-16	IM-16
SDE-40	.87***	.32***
IM-40	.28***	.84***
Self-esteem	.53***	.12***
Narcissism	.15***	-.03

Note. The values are taken from the Hart et al., (2015) paper. BIDR = Balanced Inventory of Desirable Responding; IM = impression management; SDE = self-deceptive enhancement. * $p < .05$. ** $p < .005$. *** $p < .001$.

Table 6 shows that SDE and IM had differential relations with BIDR 40 subscales and self-enhancement measures. SDE-16 and IM-16 strongly correlated with SDE-40 and IM-40, respectively. Additionally, the correlation between SDE-16 and IM-40 was weak, implying appropriate divergence of these subscales. In relation to self-enhancement measures, SDE-16 correlated strongly with self-esteem while IM-16 demonstrated weak correlation with self-esteem. SDE-16 was the only subscale that showed a significant yet weak correlation with narcissism. These patterns were mirrored by SDE-40 (self-esteem:

r =.54, *p* < 0.001; narcissism: *r* =.18, *p* < 0.001) and IM-40 (self-esteem: *r* =.09, *p* < 0.001; narcissism: *r* = -.02, *p* > 0.05). Therefore, item refinement preserved the validity of the original BIDR subscales.

Study two: ***Assessing the validity of the BIDR-16.*** The second study explored the independent validity of BIDR-16 in a different sample of online volunteers (487 females and 113 males). The study found the two-factor model of 16-item BIDR had a better fit (GFI = .92, CFI = .90, RMSEA = .07 and SRMR = .06) than the model constructed in study one. Again, retaining two dimensions improved the model fit as compared to a simple one-factor model.

Study three: ***Assessing the test-retest reliability of the BIDR-16.*** To define the independent test-retest reliability of BIDR-16, the researchers administered the psychometric scale to a sample of 352 UK students over a 2-week period. The results suggested good test-retest reliability of both subscales (SDE: *r* = .79, *p* <.001; IM: *r* = 0.74 *p* <.001). This is comparable to the original 40-item BIDR.

Study four: Assessing the external validity of the BIDR-16. Study 4 tested the external validity of BIDR-16 against the MCSDS SDR scale, self-enhancement indices (RSES, NPI-40), and personality traits (Ten-Item Personality Inventory) in an online survey of adults (588 students, 120 non-students) mainly from the UK and US (*n* = 704). The study showed stronger relationships of SDE with MCSDS, Self-Enhancement, Extraversion, and Emotional Stability as compared to IM. In contrast, IM correlated and more strongly with Agreeableness and negatively with Narcissism. SDE and IM were equally related to Conscientiousness (See table 7).

Table 7

Correlations between BIDR-16, SDE and IM, and existing scales in study 4.

Scale	SDE	IM
MCSDS	.53***	.32***
Self-esteem	.18***	.52***
Narcissism	-.18***	.26***
B-HSM	.10*	.26***
Extraversion	-.06	.16***
Emotional Stability	.26***	.45***
Openness	.04	.18***
Conscientiousness	.20***	.25***
Agreeableness	.37***	.13**

Note. The values are taken from the Hart et al., (2015) paper. BIDR = Balanced Inventory of Desirable Responding; SDE = self-deceptive enhancement; IM = impression management; MCSDS = Marlowe-Crowne Social Desirability Short; B-HSM = Brief How-I-See-Myself.

*$P < .05$. *$p < .005$. ***$p < .001$.

Limitations of the BIDR-16.

Several criticisms of the BIDR-16 are raised. In relation to study one, GFI and CFI was used to assess model fit. GFI is substantially affected by the sample size and the size of factor loading, increasing type I error and model misspecification, and is consequently no longer recommended (Sharma, et al., 2005). Additionally, the researchers used the cut-off point of 0.90 for CFI. This older cut off is lower than the recommended 0.95 value and may result in higher type II error rates (Hu & Bentler, 1999). Nevertheless, the researchers also used RMSEA and SRMR, which are robust against variations in

sample size and model misspecification. Despite satisfactory RMSEA and SRMR values and improvement in CFI of the refined 16 item scale, the CFI was still below the acceptable level (0.88 vs 0.95), limiting the fitting of this shortened two-factor model of SDR.

Furthermore, study one had issues with internal consistency[7]. Specifically, according to the Cronbach's alpha coefficient (Cronbach, 1951) the SDE subscale reliability was questionable (Alpha 0.63 to 0.66) while the reliability of the IM subscale was in the lower band of acceptable range (Alpha 0.71 to 0.74). However, it is worth noting that some statisticians contend that for scales used in human dimension research, it is adequate to use Cronbach alpha values between .65-.80 (DeVellis, 2003; Green et al., 1977; Spector, 1992; Vaske, 2008). The last point arguably supports the use of the use of the BIDR-16 (Hart et al., 2015).

Another criticism of study one is the study population consisted of a sample of adult volunteers who were recruited online and had an imbalance of female over male participants. This limits the applicability/generalisability of the psychometric to other populations (e.g., criminal justice professionals). Nor did the authors specify the self-enhancement and self-esteem measures used in the study.

In relation to study two, a drawback is that the type II error rate was high (CFI = 0.90 vs 0.95). Additionally, the internal consistency of the items clustering around each factor was not reported. A criticism of study three is that the study population was dissimilar to previous studies and no measures of validity were reported.

[7] internal consistency - refers to how well items on the same scale correlate with each other. If the correlation is high, this tells us that the items are essentially measuring the same thing.

General discussion

The BIDR-16 was developed to improve the practicality of the original BIDR-40 whilst maintaining the dimensionality, validity, and reliability of the original scale. The four studies indicated that BIDR-16 can capture two dimensions of the SDR (i.e., SDE and IM) with good reliability and relatively acceptable validity in various groups of population. The BIDR-16 can therefore be a reasonable substitute for the original long version of BIDR.

Support for the BIDR-16. A study by Bobbio and Manganelli (2011) supports the Hart et al. BIDR-16 measure. These researchers carried out three sequential studies in different groups of Italian adults to generate a shorter version of Paulhus' 40-item 6-point Likert-type response scale BIDR 6. In the first study, a translated version of the original 40-item BIDR was administered to non-student volunteers (n = 706). The scores were analysed by CFA. A two-factor model with 16 items derived from the 40-item BIDR showed good data fit ($\chi^2(103) = 226.27$, $p < 0.001$, $\chi^2/df = 2.20$, CFI = .95, RMSEA = .041, SRMR = .045). The entire scale ($\alpha = 0.71$) including its subscales of SDE ($\alpha = .70$) and IM ($\alpha = 0.74$) had acceptable reliability. There were strong correlations between the original scale and the short version (general score: $r = 0.80$; SDE: $r = 0.78$; IM: $r = 0.79$).

The two-factor model was subsequently tested in a group of university students (n = 150). The 16-item model demonstrated acceptable goodness of fit ($\chi^2(103) = 141.69$, $p < .001$, $\chi^2/df = 1.38$, CFI = .92, RMSEA = .05, SRMR = .07) and internal consistency (SDE $\alpha = 0.74$; IM $\alpha = 0.70$; entire scale $\alpha = 0.71$). The total BIDR-16 score (0.80) and its subscales, SDE ($r = 0.78$) and IM (0.80), correlated strongly with the corresponding scores on BIDR-40. This confirmed the structural validity and reliability of the shorter 16-item BIDR scale in adults.

In their second study, Bobbio and Manganelli (2011) tested the psychometric properties and factorial structure of their previously developed 16-indicator BIDR using psychology students. Participants completed an online survey containing a 5-point response 16-item BIDR. The CFA showed a suitable two-factor model fit $\chi^2(103) = 262.36$, $p < .001$, $\chi^2/df = 2.55$; CFI = .93, RMSEA = .047, SRMR = .047). However, the internal consistency of the total scale and its subscales was marginally acceptable (SDE $\alpha = .69$; IM $\alpha = .73$; entire scale $\alpha = .69$), indicating that the Web-based BIDR may have lower validity and reliability than a traditionally administered scale.

The third study explored the external validity of a 6-point Likert-type 16-item BIDR against the measures of internal locus of control, general self-efficacy, work alienation, hopefulness, and the MCSD scale in adult workers ($n = 286$). The goodness-of-fit of this 16-indicator BIDR scale (χ^2 (103) = 184.38, $p < .001$, $\chi^2/df = 1.79$; CFI = .93, RMSEA = .05, SRMR = .06) and its reliability (SDE $\alpha = .69$; IM $\alpha = .77$) were acceptable. In the correlational analysis, SDE showed a significant relationship with indicators of self-esteem (to the internal locus of control, general self-efficacy, and hopefulness) whilst IM was strongly related to MCSD and negatively to alienation. This supported the discriminant validity of the BIDR short version, and indirectly supports that the development and use of a short version of the BIDR.

However, the Bobbio and Manganelli study was derived from a translated version of BIDR-40. This may affect the accuracy and precision of this psychometric. Additionally, none of their studies explored test-retest reliability of the 16-indicator BIDR short version. Moreover, it should be noted that these studies predated the development of BIDR-16, and therefore, it does not directly support the validity and reliability of this psychometric.

The use of a convenience sample may also limit the accuracy and generalisability of the findings. Limited generalisability also applies to the scoring scale (5 and 6-point responses) of the BIDR.

Therefore, the factorial structure and reliability of the 16-item version may not be equal to a 7-point scored scale and further research is required to determine if the psychometric properties of their 16-item BIDR are comparable. Consequently, the utility of their short version should also be evaluated more widely, including with adults with different sociodemographic features. There was a lack of studies published on SDR measurement since the study by Hart et al. (2015). Given the lack of independent validation studies, the reported validity and reliability for BIDR-16 should be interpreted cautiously.

Questioning the usefulness SDR measures. Some studies have found the criterion validity of personality scales substantially reduced when participants are instructed to fake responses (Douglas et al., 1996; Holden, 2007). However, research has also found that using SDR scores to adjust personality trait scores lowers test-criterion relationships and the predictive validity[8] of the measure, thereby rejecting the assumption that SDR scores act as a statistical suppressor (cf., Dilchert & One, 2011; Paunonen, & LeBel, 2011; Reeder & Ryan, 2011). Similarly, when IM has been used as a moderator, it has not moderated the relationship between personality and the criterion (Holden, 2008); and the predictive validity of personality scales is not significantly affected when the influence of SDR is removed (Li & Bagger, 2006). Consequently, some researchers have concluded SDR scales are insensitive to respondents who distort their ratings and are imperfect proxies of faking (Holden, 2008).

Importantly, evidence indicates that social desirability may represent a more significant construct than simply response style. Factor analytic research by Uziel (2010) found that IM was related to adaptive social behaviours rather than simply self-presentation bias. Similarly, the finding that SDR scales

[8] predictive validity – is the degree to which test scores accurately predict scores on a criterion measure

overlapped with socially desirable personality traits (Ones et al., 1996) including Emotional Stability and Conscientiousness (Li & Bagger, 2006), makes it difficult to determine whether individuals who score high on SDR scales are simply virtuous or deceitful.

Another issue relates to what happens to the psychometrics of individuals suspected to be impacted by SDR. The practice of adjusting the scores of individuals suspected of faking or even removing them from the selection process, may be problematic as it disregards the possibility of false positives. It may be better to cautiously interpret the suspected SDR score.

Conclusion

This critique reviewed the BIDR-16 psychometric developed by Hart et al. (2015). The BIDR-16 is a short version of the 40-item BIDR two-factor model of SDR developed by Paulhus (1984, 1998). Overall, a review of the psychometric properties of the original 40-item BIDR has shown it to have satisfactory internal consistency, test-retest reliability, and discriminant and convergent validity (Paulhus, 1994). The BIDR-16 was developed to improve on the practicality of the original 40-item BIDR, whilst retaining its two-factor model, internal consistency, temporal stability, and discriminant and convergent validity. CFA by Hart and colleagues supported the construct of SDR into two factors and found a reasonable fit for a 16-item BIDR model (i.e., eight items per factor). A second study with an independent sample supported the findings from the first study. Temporal consistency was also found. However, study four found some overlap of SDE and IM with external correlates, which questioned whether the BIDR subscales of SDE and IM accurately represented SDR and whether SDR is an aspect of personality rather than a response style. Consequently, although the construct validity and internal consistency of the items is not optimal and warrants further refinement and supplementary studies, the overall practicality of the BIDR-16 makes it an attractive substitute for the longer BIDR.

Chapter 4: The influence of defendant and victim race on Criminal Justice Workers' decisions on guilt, punishment, and blame.

Introduction

Chapter one opened with the government statistics that Black people particularly are disproportionately overrepresented in all areas of the England and Wales CJS, compared to their representation in the general population and against their White counterparts (Lammy, 2017). Contrastingly, Asian people are generally underrepresented in comparison to their representation in the general population. The disproportionality figures have raised questions around whether the CJS in England and Wales is racially biased against certain racial groups, or whether the figures are a true reflection of the amount and type of crime committed by certain racial groups (Lammy, 2017).

Jury and juror decision making is an area of the CJS that has been extensively researched, although less so in England and Wales than elsewhere. The evidence from the research has been mixed. Some research has found juries and jurors were not racially biased (Ewanation & Maeder, 2018), others that jurors were racially biased (Maeder et al., 2015 study 2), some that racial bias was based on whether the defendant or victim's race matched the jurors' (FosterLee et al., 2006), others that racial bias occurred when the defendant's race was congruent with the type of crime (Gordon et al., 1998). The picture is complex, and the reasons given for the discrepancies in findings have ranged from cross cultural differences to the differences reflecting changes in attitudes over time (e.g., Devine & Caughlin, 2015). The conflicting findings have led to several meta-analyses, which have also provided mixed findings. The systematic review in the current thesis was conducted as previous reviews were dated and attitudes might have changed over time. However, the findings from the current systematic review were also mixed, yet importantly reinforced a lack of research in England and Wales, a lack of research on decisions concerning Asian defendants and victims; an overuse of students as participants; and that the traditional use of self-report measures raises in simulated research the possibility of socially desirable responding (SDR) and risk of bias amongst participants. The subsequent critique of

the BIDR-16 psychometric test has highlighted that it could be used to detect SDR, which if present threatens the validity of study results.

Research rationale and aims.

The rationale for this study was the gap in research using criminal justice workers as participants, particularly criminal justice workers in England and Wales. As criminal justice workers are involved all stages of the criminal justice process, it was deemed important that their attitudes be investigated to explore the possibility of racial bias in their judgements of defendants and victims. It is believed that this research will add to the literature on the decision-making processes of criminal justice workers.

The study put the criminal justice worker (CJW) in the role of a juror. Defendant and victim race were experimentally manipulated whilst keeping the other variables constant to assess influence on participants' judgements and decisions. Note, most participants involved will be exempt from sitting on a jury. However, as CJWs are usually trained to have an increased understanding of the impact of their decisions and take the task more seriously than non-criminal justice workers. Additionally, analysis was also done by professional group. As mentioned earlier, the overrepresentation of certain BAME groups at various stages of the UK CJS (see Lammy, 2017), has raised questions on whether there are specific professional groups within the CJS are racially bias. There is a lack of studies examining racial bias amongst criminal justice professionals and professions within the England and Wales. The systematic review, found that factors other than defendant/victim's race may influence mock jurors' decisions on guilt, blame or sentence. One potentially important factor which has been underexplored, is the criminal justice profession. Where studies have been conducted with CJWs and specific CJPs, these have tended to be conducted in the United States (e.g., Graham & Lowery, 2004). The lack of studies conducted in England and Wales with CJWs and different criminal justice professional groups,

highlights a gap in research. Knowing whether there are certain professions that are racially biased would allow those professions to be targeted to eliminate racial bias.

The primary hypotheses in the current thesis was that participants judgements on guilt, blame and punishment would be affected by:

[1] Defendant's race

[2] victim's race.

[3] the interaction between the race of the defendant and victim

The secondary hypotheses in the current thesis was that participants judgements on guilt, blame and punishment would be affected by:

[1] criminal justice profession

[2] the interaction between profession and defendant or victim's race

Method

Participants

A total of 399 criminal justice workers (CJWs) (49% male, 51% female) aged 21 to 71 years (*M* = 41.1, *SD* = 11.6) were recruited online over 283 days between 05.05.2018 and 10.03.2019 from 18 youth justice services (YJSs) across England and Wales, and from the social media platform LinkedIn. Other CJWs were recruited online as ethical approval for the research using CJWs employed by His Majesty's Prison and Probation Service was declined. Participants were mostly White (81.7%) and working in police (30.8%) or probation (30.3%) services (table 8). Note, overall, the number of CJWs that participated in the current study was higher than their representation in several areas within the

criminal justice system workforce. Specifically, 18.3% of CJWs who participated in this study were from the BAME population (i.e., Asian 2.8%; Black 11%; Other (which included mixed black and White, Black, and Asian, White, and Asian) 4.5%). For example, government workforce figures released in 2019 report that out of those whose ethnicity was known, police officers comprised 6.9% BAME people and 93.1% white people, compared to 15.4% BAME and 84.6 White people in the current study; prison officers comprised 7.7% BAME people and 92.3% white people, compared to 16% BAME and 84% White people in the current study; probation officers comprised 15.6% of BAME people compared with 84.4% White people; court judges comprised 7.4% BAME people and 92.6% white people, compared to 0% BAME and White people in the current study); and tribunal judges comprised 11.1% and 88.9% white people, compared to 0% BAME and White people in the current study) (see UK Government, 2021). Similarly, figures for the ethnic breakdown of social workers found that 12.8% identified as Black, and 70.4% as White (see UK Government, 2021) compared to 11.8% BAME and 88.2% White Youth Justice/ Social Workers in the current study.

Thus, comparatively, compared several areas of the criminal justice workforce, the current study has a higher proportion of BAME CJWs and lower number of White CJWs in its sample, than there is in the actual criminal justice workforce.

Table 8

Demographic characteristics

	Frequency	Percentage
Gender		
Male	196	49.1
Female	203	50.9
Ethnicity		
Asian	11	2.8
Black	44	11
White	326	81.7
Other*	18	4.5
Profession	50	12.5
Prison worker	123	30.8
Police worker	121	30.8
Probation worker	83	20.8
Youth justice/social worker		
Other*	22	5.5

Note. Other* category of ethnicity included mixed Black and White, mixed Black and Asian, and mixed White and Asian; other category of profession included interventions facilitator, forensic mental health nurse, restorative justice officer, offender support worker, and forensic psychologist.

The inclusion criteria were that participants must identify as one of the following professional groups: probation officers; youth justice case managers; social workers; police officers; other criminal justice profession (e.g., forensic psychologists); and to have recent contribution to legal decisions regarding

offenders in England and Wales. Criminal justice workers located outside of England and Wales were excluded.

Questionnaires were distributed to 635 criminal justice workers. Of these, 209 (32%) were not returned, and 27 questionnaires (4%) were excluded - 25 for withholding consent and 2 for not meeting inclusion criteria because they were completed by criminal justice workers located outside of England and Wales (Scotland).

The study received ethical approval from the University of Nottingham's Faculty of Medicine & Health Sciences Research Ethics Committee (ethics reference No: 289-1803. See Appendix F).

Design

This cross-sectional between groups study employed a 3x3 factorial design combining two independent variables (defendant and victim race) at three levels (Black, White, or Asian), with three dependent variables (defendant guilt, blame and type of sentence). Participants were allocated to one of nine experimental conditions based on the combination of victim and defendant race.

Power

Spencer (2003) investigated racial bias in mock-juror decisions by students and found the effect size of a significant interaction between victim and defendant race to be 0.119 (η^2), a small to medium effect size. Using Cohen (1988) this equates to an effect size in f^2 of 0.135. Power analysis was conducted for Multivariate Analysis of Variance (MANOVA) in G*Power (Faul *et al.*, 2013) using an α value of 0.05, a β value of 0.80, and an effect size in f^2 of 0.135. The analysis indicated a minimum sample size of 88 participants was needed.

Measures

Questionnaires were administered to each participant at a single time point via Bristol Online Surveys platform (Bristol Online Survey, 2018), in the following order (see also appendix G):

Participant demographic questionnaire

The non-standardised questionnaire designed for this study collated demographic information on age, gender, ethnicity, and CJP (See Appendix H).

Balanced Inventory of Desirable Responding – 16 (BIDR-16; Hart, Ritchie, Hepper and Gebauer, 2015) (See appendix E).

The BIDR-16 is a 16-item Likert-type self-report measure of socially desirable responding (SDR), developed to be administered alongside other self-report measures. Participants are required to rate the extent to which each statement applies to them using an 8-point scale: 1 (totally agree) – 8 (totally disagree). It has two subscales: Self-Deception Enhancement (SDE - denial of undesirable self-descriptions and inflation of personal characteristics) and Impression Management (IM - an inclination to create favourable impressions of oneself to others). It has been used to assess for the likelihood of fabricated responses (Paulhus, et al. 1995). SDE and IM have scores ranging between 8 and 64 and have 8 items each.

The possibility of SDR in participants' decisions was determined by testing the relationship of BIDR 16 (IM and/or SDE) scores with measures of guilt, blame and punishment. The presence of a significant relationship would indicate SDR. This effect of SDR could then be controlled for by including the significantly related subscales as covariates in the analytical models (Hart et al., 2015). The BIDR-16 has been found to be reliable and valid psychometric measure in the general population in the UK (Hart et al., 2015), though it has not been evaluated in CJPs. The psychometric evaluation of the BIDR-16 was

the subject of the previous chapter.

Court case vignette (see appendix I)

Participants were presented with a 270-word written vignette briefly describing a fictional court case describing an assault. The vignette was presented as a verbal exchange between prosecuting lawyer and defence lawyer and was broken into three sections: [1] court case summary, [2] the prosecuting lawyer's summary, and [3] the defence lawyer's summary.

Vignettes have frequently been used in the methodology of simulated jury/juror studies. They allow researchers to simultaneously explore several themes, efficiently review participants' interpretation of events, and make recommendation for best practice (Finch, 1987). Vignettes may be less intrusive than direct discussions at encouraging participants to discuss sensitive topics (Finch, 1987), and may help to reduce socially desirable responding and experimenter bias (Hughes & Huby, 2004; Schoenberg & Ravdal, 2000). Vignettes have been acknowledged as an effective tool to discover the opinions, attitudes, and perceptions of participants (Nygren & Oltedal, 2015).

A written vignette was chosen because of its cost effectiveness and ease to produce compared to other medium such as video vignettes; and some evidence indicating the responses they elicit from participants (e.g., blame attribution for a physical assault) are not significantly dissimilar to other vignette methods (Lanza et al., 1997). The length of written vignettes in research varies, with some research showing a length of between 150 – 300 words is sufficient (Barter & Renold, 2000; Finch, 1987; Hughes & Huby, 2004). The length of the vignette was chosen to try to maximise participation, as vignette research has found that compared to longer vignettes, shorter vignettes can help increase task completion rates, reduce the likelihood of fatigue and careless responding on follow-up

questionnaires (Lawrie et al., 1998; Nosanchuk, 1972). In the current study the length of the vignette is 270 words, which was considered adequate to keep the participants interest and engage them with the scenario (Barter & Renold, 1999; Bradury-Jones et al., 2014).

There is some evidence to show that there are race based crime stereotypes (Jones & Kaplan, 2003; Skorinko & Spellman, 2013). There is also evidence that people with non-white sounding names are discriminated against in areas including employment, the renting of property and discriminatory victim blame (e.g., Arai & Skogman-Thoursie, 2009; Bertrand & Mullainathan, 2004; Clark et al., 2013). Such findings highlight the difficulties in identifying a race neutral offence and name and demonstrates the need to control the influence of both factors. Consequently, in the current thesis the crime used in the description has been identified as stereotypical for Black male defendants (i.e., physical assault) (Sunnafrank & Fontes, 1983) whereas the surnames (Taylor and Smith) used for the defendant and victim were names common amongst the White British population (Top British Surnames, n.d.). This approach was taken to avoid the names of defendant and victim uncontrolledly activating stereotypical race crime effects, and confounding investigation of the variables of interest: the race of defendant and victim. These measures can also be viewed as salient ways to introduce race to the participant. This method falls in line with the mock juror paradigm model of research introduced by Landy and Aronson (1969), who stated that only the (independent) variables of interest should be manipulated. This means manipulating the race of the defendant and victim in the current studies whilst keeping the other facts of the case constant, like their names.

Post-trial questionnaire (see appendix J).

This was a set of 10 non-standardised questions about the vignette adapted from Gordon (1993), asking participants about their opinions regarding guilt, blame and punishment for the accused in the vignette. Responses were rated categorically.

Variables

Outcome variables

Guilt was measured on a 7-point scale response to the following question: "Based on what you have read and the instructions that you have been provided, how likely do you find/rate the defendant guilty of grievous bodily harm?"

Blame was measured on a 5-point scale response to the following question: "How much blame do you think the victim should take for the offence?"

Sentencing decisions were categorised as the prison sentence, community sentence, other punishment, or no punishment.

Independent variables

The defendant's and victim's race and the type of CJP were defined nominally.

Control variable

SDR was measured by total scores on the BIDR-16 questionnaire, ranging from 16 to 112.

The SDE and IM subscale scores on the BIDR-16 were also defined continuously, ranging from 8 to 56 each.

Procedure

The Participant Information Sheet (PIS) and Recruitment Poster (Appendices K & L) were emailed to all Youth Justice Services in England and Wales. In addition, the recruitment poster was placed on the social media platform LinkedIn. The PIS was subsequently emailed to those who expressed their interest, along with the research aims, participants' demographics form and confidentiality clause.

Participants were also informed of the procedures and that there was no obligation to participate, and they could withdraw at any stage. The researcher's contact details were also included so that the participants could ask questions.

The participants were randomly allocated to one of nine conditions using the Select Cases function of SPSS Version 24, applying Arifin's (2012) procedure for pseudo-block random allocation. (See appendix M for experimental conditions). They were then emailed a link to an online questionnaire for their condition, with a unique participant number to match them to their completed survey questionnaires. Participants were asked to confirm consent at the start of the online questionnaire (see Appendix N).

Each participant was then presented with the study materials in the order given above. The entire protocol was estimated to take 15 minutes.

On completing the survey questionnaires, participants read the Debrief Sheet (see Appendix O), were provided with the opportunity to email the researcher any questions and reminded of their right to withdraw.

Statistical analysis

The data were statistically analysed using SPSS version 24. Data distribution was checked using the Shapiro-Wilk test of normality and visual inspection of histograms. Where distributions were non-normal, non-parametric tests were selected. To control for the inflation in type I error following pairwise comparisons, the α value for significance tests was set at $p < 0.01$.

Because the BIDR's IM subscale was skewed, Spearman's correlation test was selected to assess the relationship between IM, guilt, and blame. Since total BIDR and SDE scores were normally distributed, their correlations with guilt and blame were examined using Pearson's test. To test the effect of SDR on type of sentence, multinominal logistic regression was used. To measure the main and interaction

effects of the defendant's and victim's race on guilt and blame, a factorial Multivariate analysis of variance (MANOVA) was conducted. Pairwise comparisons for significant interactions were undertaken when required. Chi-square tests with contingency tables were undertaken to study the association of defendant race, victim race, and criminal justice professions with the sentencing decision categories. When the assumption of having <20% of cells with a count <5 within the crosstabulation was not met, the likelihood ratio (LR) was used as an alternative to the Pearson's Chi-square test. Cramer's V was used as the measure of effect size for the non 2 × 2 contingency tables. Significant findings were explored further by post-hoc analysis to identify subgroups of CJP which differed significantly in the sentence severity they ascribed to the defendant.

Results

The rated guilt and blame as well as the SDR scores are presented in table 9.

Table 9

Descriptive characteristics of the entire study population

	Mean	*SD*	Minimum	Maximum
Self-Deceptive Enhancement	37.49	6.74	20.00	55.00
Impression Management	39.00	7.64	14.00	56.00
Total BIDR-16	79.68	12.26	35.00	109.00
Blame	1.19	.66	1.00	5.00
Guilt	3.82	1.57	1.00	7.00

Note. Values are presented as mean and *SD.*

For all variables *N* = 399

Possibility of Socially desirable responding (SDR)

There was no significant relationship between participants' judgement on guilt and blame and the IM and SDE scores (Table 10), suggesting that it was unlikely that SDR was present.

Table 10

Correlation of socially desirable responding (as measured by BIDR-16) with guilt and blame.

	Guilt		Blame	
BIDR-16	*r*	*p* value	*r*	*p* value
SDE	.07	.16	.05	.81
IM	-.02	.66	-.06	.22
Total SDR	.02	.66	-.02	.73

Note. For all variables *N* = 399

SDE: self-deceptive enhancement; IM: impression management; SDR: socially desirable responding

Likewise, multinomial logistic regression using the Wald's statistic did not show any significant effect of total SDR on measures of punishment. Neither SDE (for community punishment, $W(1) = .04$, $p = 0.85$; for prison, $W(1) = 3.38$, $p = 0.07$; for other punishment, $W(1) = .86$, $p = .35$) nor IM (for community punishment, $W(1) = .44$, $p = .50$; for prison, $W(1) = 1.36$, $p = .24$; for other punishment, $W(1) = .11$, $p = 0.74$) were significantly associated with type of sentence.

Criminal justice workers' judgements on guilt and blame by defendant and victim's race.

Table 11 shows the variation of perceived guilt or attributable blame by defendant's and/or victim's race.

Table 11

Perceived guilt and attributed blame by defendant and victim's race

	Guilt		Blame	
Defendant race	Mean	*SD*	Mean	*SD*
Black	3.72	1.60	1.21	.62
White	3.92	1.52	1.14	.57
Asian	3.83	1.61	1.22	.78
Victim's race				
Black	3.95	1.55	1.23	.72
White	3.74	1.65	1.11	.50
Asian	3.79	1.53	1.23	.73
Defendant vs victim's race				
Black - Black	3.60	1.60	1.27	.65
Black - White	3.74	1.71	1.14	.52
Black - Asian	3.80	1.53	1.22	.69
White – Black	4.07	1.45	1.22	.76
White – White	3.79	1.61	1.07	.34
White – Asian	3.91	1.52	1.14	.51
Asian - Black	4.19	1.56	1.19	.74
Asian - White	3.70	1.67	1.13	.62
Asian - Asian	3.64	1.57	1.33	.93

SD: standard deviation

Additionally, factorial MANOVA found that the defendant's race had no significant effect on perceived guilt, F (2, 390) = .57, p = .56 or attributed blame, F (2, 390) = .50, p = .60. Factorial

MANOVA also indicated that victim's race had no significant on the perceived guilt, F (2, 390) = .64, p = .53 or attributed blame, F (2, 390) = 1.32, p = .27. Furthermore, there were no significant differences in perceived guilt, F (4, 390) = .74, p = .56 or attributed blame, F (4, 390) = .39, p = .82 across the combination of defendant and victim's race (Appendix P)

Criminal justice workers' sentencing decisions by defendant and victim's race.

Overall, a larger proportion of CJWs suggested a prison sentence when the defendant was described as Asian (48.1%) vs White (38.6%) or Black (37.3%). Also, more CJWs suggested a prison sentence when the victim was described as either Asian (43.7%) or Black (46.2%) vs White (34.1%) (Table 12).

However, the Chi-square test (table 12) did not find a significant association between sentencing decisions and defendant's race, $\chi^2(6) = 12.86$, p = .04, or victim's race, $\chi^2(6) = 5.04$, p = .53.

Table 12

Suggested sentence by defendant and victim's race.

	Community sentence		Prison		Other		No punishment	
	Count	%	Count	%	Count	%	Count	%
Defendant's race								
Black	28	20.9%	50	37.3%	25	18.7%	31	23.1%
White	27	20.5%	51	38.6%	27	20.5%	27	20.5%
Asian	27	20.3%	64	48.1%	18	13.5%	24	18.0%
Victim's race								
Black	31	23.5%	61	46.2%	15	11.4%	25	18.9%
White	23	17.4%	45	34.1%	28	21.2%	36	27.3%
Asian	28	20.7%	59	43.7%	27	20.0%	21	15.6%

Table 13.

Perceived guilt and attributed blame by CJWs' professions

Criminal justice professions	Guilt			Blame		
	Mean	*SD*	*N*	Mean	SD	N
Other criminal justice professional	3.59	1.593	22	1.36	.902	22
Police Worker	3.93	1.825	123	1.10	.413	123
Prison Worker	3.56	1.500	50	1.32	.794	50
Probation Worker	4.17	1.315	121	1.22	.769	121
Youth Justice /Social Work	3.39	1.464	83	1.16	.614	83

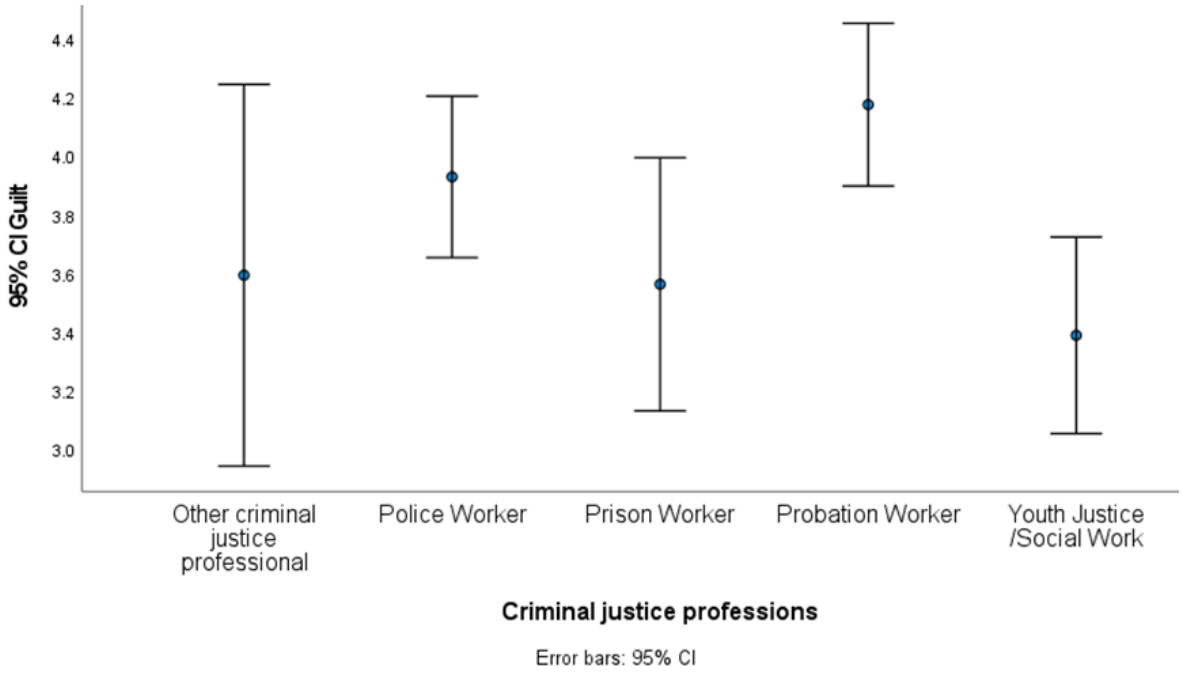
SD: standard deviation

Participants' judgements on guilt, blame by profession.

The levels of perceived guilt and attributed blame by different CJPs are presented in table 13. According to MANOVA, CJP had a small but significant effect on perceived guilt among the mock jurors, F (*4*) = 3.79, p = .005, η_p^2 = .04. Based on the Bonferroni corrected post-hoc analysis, probation workers attributed a significantly higher level of guilt to the defendant compared with youth justice/ social workers, *MD* = .79, p = .004 (figure 2).

Figure 2.

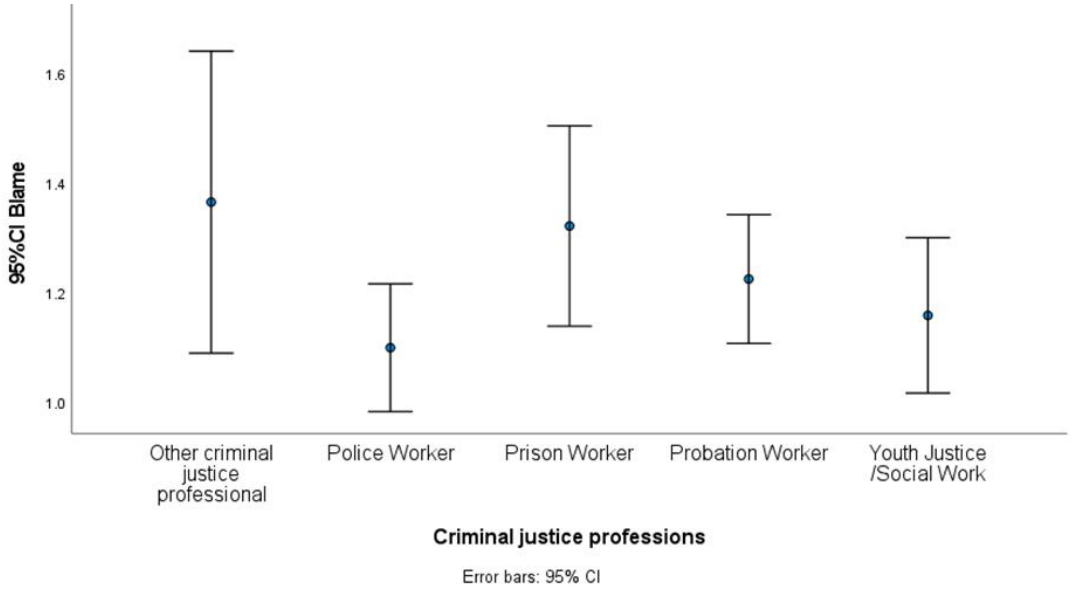
Perceived mean guilt by CJWs' profession



However, as shown in figure 3, profession did not have a significant effect on level of blame attributed to the defendant or victim race, F (4) = 1.60, p = .17.

Figure 3.

Attributed blame across CJWs' professions



Participants' sentencing decisions by profession.

The Chi-square test showed that the suggested sentences varied significantly across criminal justice professions ($\chi^2(12) = 54.41$, $p < .001$). Youth justice/social workers chose either no sentence (32.5%) or community sentence (36.1%), whereas most other professionals (40-56.1%) suggested the prison sentence. In comparison, a minority of youth justice/social workers suggested prison (18.1%) or "other sentence" (13.3%). Youth justice workers/social workers (32.5%) followed by prison workers (26%) were the professionals who were most likely to suggest "no sentence" (Table 14).

Table 14

Suggested sentencing across criminal justice professions.

Professions	Type of Sentence			
	No sentence	Prison	Community Sentence	Other
Other	18.2%	45.5%	9.1%	27.3%
Police	19.5%	56.1%	5.7%	18.7%
Prison	26.0%	40.0%	18.0%	16.0%
Probation	11.6%	42.1%	28.1%	18.2%
Youth Justice /Social Work	32.5%	18.1%	36.1%	13.3%

Participants' judgements on guilt, blame and punishment by profession and defendant race.

No significant interaction was found between defendant race and profession on blame (F (14, 384) = .91, p = .55) or guilt (F (14, 384) = 1.91, p = .02) (Appendix Q). However, profession and defendant race had large and significant interactive effects (Table 15) on the type of sentence for the defendants described as Black, $\chi^2(12) = 31.28$, V = .28, p = .002; White, LR (12) = 31.39, V = .24, p = .02; Asian, LR (12) = 28.36, V= .27, p = .005. As presented in Table 15, youth justice/social workers were most likely (42.3%) to choose community punishment and least likely to choose prison (19.2%) for a defendant described as Asian compared to other professions. In contrast, prison workers were most likely to choose no punishment for defendants described as Asian (41.2%). For a defendant described as White, police workers had the highest likelihood of choosing prison (58.3%), with no police worker selecting the community punishment option (0%). For a defendant described as Black, youth justice/social workers were the most lenient group of CJWs, with the lowest likelihood of imprisonment decisions (15.2%) and the highest likelihood of choosing no sentence (39.4%) compared to other groups of CJWs.

Sentencing decisions did not vary significantly by defendant's race within each group of criminal justice professions (Appendix P).

Table 15

Suggested sentence across criminal justice professions by defendant's race.

Defendant's Race	Profession	Type of sentence			
		No Sentence	Prison	Community Sentence	Other
Black	Other (*n* = 10)	2.3%	50%	0%	40%
	Police (*n* = 42)	9.7%	47.6%	4.8%*	23.8%
	Prison (*n* = 11)	0%	45.5%	27.3%	27.3%
	Probation (*n* = 38)	18.4%	39.5%	31.6%	10.5%
	Youth justice/ social worker (*n* = 38)	39.4%*[a]	15.2%*	33.3%	12.1%
White	Other (*n* = 5)	40%	40%	20%	0%
	Police (*n* = 36)	19.4%	58.3%*	0%*	22.2%
	Prison (*n* = 22)	27.3%	36.4%	18.2%	18.2%
	Probation (*n* = 45)	11.1%	33.3%	31.1%	24.4%
	Youth justice/ social worker (*n* = 24)	29.2%	20.8%	33.3%	16.7%
Asian	Other (*n* = 7)	14.3%	42.9%	14.3%	28.6%
	Police (*n* = 46)	15.6%	62.2%	11.1%	11.1%
	Prison (*n* = 17)	41.2%*	41.2%	11.8%	5.9%
	Probation (*n* = 38)	5.3%	55.3%	21.1%	18.4%
	Youth justice/ social worker (*n* = 26)	26.9%	19.2%*	42.3%*	11.5%

Note. n indicates the number of respondents to the scenarios from each type of profession.

*The flagged significant associations indicate the professional group with significantly different decisions on no punishment or any given sentence at significance level $p < .01$

Participants' judgements on guilt, blame and punishment by profession and victim race.

No significant interaction was found between victim's race and profession for either guilt (F (14) = 1.64, p = .06) or blame (F (14) = 1.78, p = .04) (Appendix P). However, the victim's race significantly affected the association between profession and sentencing decisions (Table 16), with larger effects being observed for victims described as Black or Asian ($\chi 2(12)$ = 27.08, V = .*26, p* =.002 for Black; *LR* (12) = 35.73, V = .30, p < .001 for Asian. When the victim was described as Black, youth justice workers/social workers were more likely (43.3%) to choose 'community sentence' compared with police (4.7%), prison (17.6%), and probation workers (37.1%) (Table 15). Police workers rarely chose community punishment. For a victim described as Asian, a large proportion of youth justice/social workers chose no sentence or community punishment (66.6%) whilst police workers (58.1%), prison workers (47.1%), and probation workers (42.9%) chose prison more often. However, no significant difference in sentencing decisions was found between professions when the victim was described as White, ($\chi 2(12)$ = 20.58*, p* = .06).

Table 16

Suggested sentence across criminal justice professions by victim's race

Victim's Race	Profession	Type of sentence			
		No Sentence	Prison	Community Sentence	Other
Black	Other (*n* = 7)	14.3%	85.7%	0%	0%
	Police (*n* = 43)	20.9%	58.1%	4.7%* [a]	16.3%
	Prison (*n* = 17)	23.5%	47.1%	17.6%	11.8%
	Probation (*n* = 35)	11.4%	42.9%	37.1%	8.6%
	Youth justice/ social worker (*n* = 30)	23.3%	23.3%*	43.3%*	10%
White	Other (*n* = 10)	10%	20%	20%	50%*
	Police (*n* = 39)	10.3%	61.5%*	7.7%	20.5%
	Prison (*n* = 17)	11.8%	52.9%	23.5%	11.8%
	Probation (*n* = 40)	5%	47.5%	25%	22.5%
	Youth justice/ social worker (*n* = 29)	41.4%*	17.2%*	31%	10.3%
Asian	Other (*n* = 7)	14.3%	85.7%	0%	0%
	Police (*n* = 43)	20.9%	58.1%	4.7%*	16.3%
	Prison (*n* = 17)	23.5%	41.1%	17.6%	11.8%
	Probation (*n* = 35)	11.4%	42.9%	37.1%	8.6%
	Youth justice/ social worker (*n* = 30)	23.3%	23.3%*	43.3%*	10%

Note. n indicates the number of respondents to the scenarios from each type of profession.

*The flagged significant associations indicate the professional group with significantly different decisions on no punishment or any given sentence at significance level $p < .01$

Discussion

This is the first study in the UK which explored the possible influence of racial bias in criminal justice workers using the mock-juror paradigm.

Racial bias in judgements on guilt, blame and punishment.

Findings from the present study did not support the hypotheses that the race of the defendant, victim or their combinations would influence participants' decisions on guilt, attribution of blame and the type of sentence.

These findings are consistent with a study in England and Wales by Thomas (2010) and a meta-analysis conducted in the US by Mazella and Feingold (1994), which found no evidence of racial bias in sentencing decisions. However, the finding is inconsistent with the US study by Toobin (1996), which found that Black mock jurors gave greater leniency toward Black defendants.

The present study found no evidence for the influence of race-crime congruency on participants' attribution of blame in the setting of Black-race stereotypical crime. This finding receives mixed support from the literature. Simulated research in England and Wales by Thomas (2010) found that White jurors neither held a race-based crime stereotype for assault, nor treated Black defendants unfavourably. However, the finding is inconsistent with that of Sunnafrank and Fontes (1983) who reported that assault was a race-stereotypical crime for Black defendants, and Gordon (1990) who reported that defendants accused of race stereotypical crimes were considered more blameworthy, more typical offenders, and deserving of longer sentences, than defendants accused of non-race stereotypical crimes. The discrepancy in findings might reflect changes in opinions over time, or cultural differences between UK and US participants in terms of what constitutes race-stereotypical crimes and their possible influence on juridical decisions.

The present research also did not find a significant influence of victim's race on judgements concerning guilt, blame and sentencing decisions. Consistent with this, Thomas's (2010) simulated research found

that victim race did not bias jury decisions. Similarly, in relation to structured professional judgements, Munoz et al., (2021) using a sample of 137 juvenile probation officers (JPOs) from five US states, found JPOs did not show ingroup or outgroup racial bias in their overall risk judgements using the Structured Assessment of Violence Risk in Youth (SAVRY). However, the latter findings have been contradicted. For example, in relation to juror decisions, Johnson et al., (2010) found that defendants with White victims were more likely to be treated more severely than defendants whose victims were Black. The mixed findings might reflect differences in decisions made using validated structured professional judgements tools and the use unstructured professional judgement methods (e.g., juror decisions), the latter of which research suggests is more susceptible to biases (Hanson & Morton-Bourgon, 2009). However, it is worth noting that although some research has found risk assessment tools not to be racially biased (e.g., Lowder et al., 2019; Perrault et al., 2017), there is research that found the opposite – particularly those that rely on actuarial static factors (Barabas et al., 2019; cf. Holder, 2014; Larson et al., 2016; Robinson & Koepke, 2019; St John et al., 2020).

Similarly, there was no significant interaction of defendant and victim race on the above decisions. This contradicts Feild's (1979) observation that the largest racial bias occurred when defendant and victim race were mismatched. In keeping with this, US simulated and archival research show that Black defendants accused of interracial offences against White victims tend to receive severer sentences than White defendants whose victims are Black (Paternoster & Brame 2008). However, these studies were conducted some time ago, and it is possible the discrepancy in findings reflect a decline in racial bias, or possibly that racial bias has become more covert.

These inconsistencies may reflect differences in views on race stereotypical crimes between the US and UK. It might also reflect a decline in racially biased decisions over time as the findings which are inconsistent with the present study tend to be from several decades ago. Alternatively, the

inconsistencies may reflect differences in race-based crime stereotypes held by the different participant populations (i.e., undergraduate versus criminal justice workers).

The non-significant finding might reflect the influence of antibias training on CJWs' decision making. Note, anti-bias and unconscious bias training aims to reduce bias through challenging racial stereotypes, attitudes, beliefs, and behaviours; and teaching people strategies to reduce bias (cf. Carnes et al., 2015). A two-year longitudinal study by Devine et al. (2012) which delivered unconscious racial bias training to 292 US college students, found that those who completed the training expressed more antibias views and challenged racial bias and racial prejudice compared to students who did not complete the training. This reduction was found up to three months after the training. These findings have been replicated with other student samples and even cross-culturally with Chinese pre-schoolers (Forscher, 2017; Qian et al., 2019).

It was not identified whether CJWs had completed any anti bias or unconscious bias training prior to their involvement in the current empirical research. Extrapolating from student samples, it is possible that such training has similar effects with CJWs, and so could explain the lack of significant results in empirical studies of the current thesis. However, a recent systematic review by FitzGerald et al. (2019), that included 47 interventions designed to reduce racial bias amongst participants, found mixed evidence of anti/unconscious bias training being effective - with some studies showing a decrease in racial bias and others an increase in racial bias following anti/unconscious bias training.

Alternatively, the discrepancies may indicate insufficient race saliency in the crime descriptions used in the current research. Some studies have found that White jurors are less likely to be racially biased when the defendant's race is made salient (or more obvious e.g., using visual images of the defendant) (Sommers & Ellsworth, 2001), perhaps from the desire to appear unbiased (Dovidio & Gaertner, 2004).

Race could have been made more salient by increasing the number of times the race of the victim/defendant was mentioned in the vignette, or instead through use of photographs of videos.

Another possible explanation for the non-significant racial bias among participants in this study might be the overwhelming strength of the evidence against the defendant. Research has found that extra-legal factors including personality (Kaplan & Miller, 1978), gender and socioeconomic status (Dane & Wrightsman, 1983) may influence juror decision making when the evidence against the defendant is weak or inconclusive, but jurors base their decisions solely on the evidence when it is strong (Gordon, 1993). This is consistent with the liberation hypothesis (Kalven, et al., 1966), that proposes that juror biases are unlikely to influence decisions unless the evidence is so ambiguous that it neither favours the defendant or the victim.

Judgements on guilt, blame and punishment by criminal justice profession.

This research found a significant association between criminal justice profession and perceived guilt among CJWs, with probation workers suggesting a higher degree of guilt for the defendant compared with youth justice/ social workers. This finding is consistent with some US based studies. For example, Graham and Lowery (2004) found police and probation officers who were primed with words stereotypically associated with Black people, were more likely to rate offenders (with unspecified race) higher on culpability, to recommend harsher punishments, and to predict higher rates of recidivism, than officers in a neutral word condition. It is noteworthy that the current study also used a Black race-stereotypical crime scenario to prime the participants.

In contrast, CJWs' professions did not have a significant main effect on their decisions regarding the blame attributable to the victim. This is inconsistent with a study by Bridges and Steen (1998), which found that probation officers tended to attribute the causes of crime committed by Black and White

males respectively to internal and external causes. Bridges and Steen's finding has received support (Freiburger & Hilinski, 2011; Hart, 2016; Shepherd & Anthony, 2018). Possibly, the discrepancy between the findings from the current study and previous studies might reflect differences in the attribution of blame between CJPs in England and Wales versus CJPs practicing elsewhere.

Of note, sentences suggested by the participant CJWs were significantly influenced by their profession. Among CJWs, youth justice/social workers were less likely than other criminal justice professionals to suggest prison sentence. Interestingly, the association between profession and suggested sentence varied across defendant and victim race groups. For example, when the defendant was described as Asian, youth justice/ social workers were less likely to suggest prison, and prison workers were more likely to choose 'no punishment' compared with other professions. Likewise, when the victim was described as Black, youth justice /social workers were more likely and police workers were less likely to suggest community sentence compared with other professions. Also, police officers were more likely to choose prison when the defendant was described as Asian and least likely when the defendant was described as Black. The observed discrepancy between professions might be that youth justice/ social workers are more used to making decisions about juvenile defendants and/or offenders and victims, the role of which might have the tendency to take a more caring and lenient position to punishment and sentencing than professionals working with adults. The latter might explain why in the current study youth justice/ social workers tended to be more lenient in their decision compared to other professions. Specifically, different professions might deal with defendants and victims differently, which then influences their decision making.

Aside from type of profession perhaps impacting on CJPs' decisions, other factors that have been found to impact on decision making include CJPs' experience, training, and length of service (Leifker & Sample, 2011); ambiguity of the evidence (Charman et al., 2017) and whether the party is representing

the defence or prosecution (Murrie et al., 2009). Also, the CJPs' experience of dealing with the type of offence used in the current study (i.e., familiarity with case scenario), might have influenced their decisions.

Overall, the present study found no evidence of racial bias in the judgements by criminal justice workers on the attribution of blame, perception of guilt, and sentencing decision. However, participants' decisions on guilt and punishment were influenced by their profession.

Study methodology issues.

This study has several strengths and methodological concerns which are discussed below.

Recruitment and participants

Some research has found that compared to traditional recruitment methods (e.g., flyers, word of mouth), online recruitment methods improve time, cost, and provide a broader representation of hard to research groups (Whitaker et al., 2017). However, the online recruitment could have led to selection bias limiting the recruitment of a diverse population (Stanton, 1998) and consequently generalisability of the findings. Some research has found that that using internet-based methods to recruit participants can lead to an overrepresentation of participants who are educated, White and female (Benedict et al., 2019; Whitaker, 2017). Furthermore, such methods are reliant on participants having internet access, which limits the scope of participants who can take part. In relation to the use of LinkedIn in this research, it is possible that criminal justice workers who used the LinkedIn platform were demographically dissimilar to other criminal justice workers (Dillman, 2000; Wright, 2005). Additionally, since some participants did not provide their work email, their profession could not be verified.

The study also had a relatively high non-return rate (32.8%). It is unclear whether the non-return rate was due to participants just not completing it or starting and then stopping completing the questionnaire. If the reason for not completing was the participants finding out about the topic and then pulling out due the sensitivity of it, then this might have led to attrition bias (Bankhead et al., 2017). It has been proposed that attrition rates of >20% threatens the validity of the results (Schulz & Grimes, 2002). Furthermore, some research has found that people who drop out from research are demographically different from those who remain (Damen et al., 2015). The implication for the current study is that the criminal justice workers who remained in the research might have been characteristically different from those who left. Perhaps criminal justice workers with more overtly racially biased attitudes, who realised the purpose of the study were more likely to withdraw. This could have influenced the non-significant results. Note, mock juror research by Sommers and Ellsworth (2001) found that when race was made less salient, racial bias amongst mock jurors increased. Perhaps, increased race saliency might increase the likelihood of participants deciphering that the study aimed to explore the influence of race on decisions, with those who are racially biased being more likely subsequently withdraw or respond in ways to appear unbiased (Dovidio & Gaertner, 2004). Consequently, to avert the possibility of racially biased participants withdrawing from research due to deciphering the research hypothesis, it is suggested that future research make race less salient in the vignettes through reducing the amount of times race is mentioned, and/or employing photographs/video footage of defendants and victims to make race more subtle.

Use of online surveys

The study used an online survey to collect data from participants. This allowed the questionnaire to be quickly and easily distributed to a broad range of criminal justice workers across a large geographical area with minimal cost (Garton et al., 1999; Llieva et al., 2002). The method also allowed data to be

securely collected and stored. However, it is possible that participants restarted the survey once the aims/objective of the study were known to them. This might have introduced SDR. Additionally, the software programme did not allow the criminal justice workers to reread the case vignette, restricting the flexibility of reviewing the evidence and limiting the precision of decisions. This could have led to the recall bias. Furthermore, a setback with using online surveys to collect data is that we cannot know for certain that the participant completing the survey are who they say they are. In the current study, attempts were made for participants to verify who they were through use of their work emails, but this does not prevent them from passing the questionnaire on for someone else to complete. Better ways of verifying the identity of participants in online surveys is perhaps required, particularly considering more research has relied on the use of online surveys in response to the Covid-19 health pandemic, limiting the use of in-person research methods.

Materials

The study used case vignettes to explore racial bias. An advantage of vignettes is that it is considered to be a less intrusive way than direct discussions to get participants to discuss sensitive topics (Finch, 1987), which may help reduce socially desirable responding and experimenter bias (Hughes & Huby, 2004; Schoenberg & Ravdal, 2000). However, the validity of vignettes as a research tool is questionable. Vignettes, particularly in the written format, have limited ecological validity (Hughes & Huby, 2004; Paddam et al., 2010). Using written vignettes might disconnect participants from the case, leading to mismatch between their responses in a theoretical setting compared with their decisions in practice (Hughes & Huby, 2004; Paddam et al., 2010). Limited content validity of the vignettes (Torres, 2009) might also impact on participants' responses (Bradbury-Jones et al., 2014). Selecting a task more consistent with criminal justice workers' actual duties might have increased the validity and consequently the generalisability of this study.

Critics of the mock juror paradigm have also argued that the methodology artificially inflates the influence of extra-legal variables through isolating manipulated variables (Bartol & Bartol, 1994). As a result of these criticisms, the behaviours of participants in simulated studies may be different from those expressed in real work settings.

Note, if this empirical research were to be extended in the future to inform policy around jury selection and processes, then it would need to ensure that trained legal professionals (or an expert panel) are involved to ensure that the evidence presented to participants in the vignettes are legally admissible within the English and Welsh legal jurisdiction; trial transcripts rather than trial summaries are used; and the vignettes include the judge's instructions to jurors (cf. Willmott et al., 2021). It is also suggested that transcripts from a real case is used or longer vignettes, which although may increase the length of the study and the potential of participant drop out, would arguably improve the ecological validity of the study (see Willmott et al., 2021).

The influence of geographic location, training, and interracial contact

Some research using children as participants has found that increased interracial contact reduces prejudice in children; whilst other research has found that levels of racial outgroup bias by White and Asian people living in diverse neighbourhoods, decreased when level of positive racial interactions increased (Binder et al., 2009; Kros & Hewstone, 2020). Extending these finding to the field of CJWs' decisions, perhaps then, the type and quality of interracial/ interethnic experiences that CJWs have influences their decisions.

Another potential limitation of this study concerns the use of adult defendants and victims in the vignettes, the use of which might suit and be more applicable to professions (e.g., Probation Officers) that deal with adult defendants. Therefore, the task for youth justice workers could be considered less

ecologically valid and applicable to their job role. Importantly also, any observed difference might be more to do with a lack of experience in making decisions regarding adult defendants/ victims, rather than possible racial bias.

Conceptual limitation

A possible conceptual limitation of the study is that all participants were asked to comment on the type of sentence that they thought the defendant should be given. Although the responses were 'prison, community punishment, no punishment', by allowing all participants/CJWs to answer this question might have conflated the responses. It is possible that participants/CJWs believed that defendants were guilty but should receive 'no punishment'; and/or that they were 'not guilty', therefore should receive 'no punishment'. On reflection, to avoid the possibility of conflating the information gathered, it might have been better to restrict the question to be answered only by participants/ CJWs who thought the defendant was guilty.

Also, it is important to note that 'blame' was not clearly defined in this empirical research. This lack of clarity could have led to varying interpretations of blame by participants, thereby ultimately impacting on the overall results. Any future research will need to have a clear definition of what is meant by blame.

Statistical analysis

There are some statistical limitations to consider. First, regarding the use of BIDR-16 as a measure of SDR, the validity and reliability of this psychometric has not been tested in a similar population, perhaps questioning its applicability for use with criminal justice workers. Second, the survey questionnaires used to measure participants' judgements on guilt, blame and punishment were not assessed for reliability. Third, the low numbers of participants from different racial groups both within

and across professions has limited the scope of statistical analysis that could be done. For instance, it may have been interesting to explore the impact of the race of all criminal justice worker on their decisions, to explore the possibility of in-group and outgroup racial bias. However, in relation to BAME participants, the low participant numbers meant that it was not statistically justifiable to explore this. Although a lower alpha value (p <0.01) was considered for statistical significance to reduce the chance of type I error, it might have resulted in some important associations being reported as non-significant because of increased risk of type II error (Andrade, 2019; Greenland et al., 2016). Further, the utility of p-value as an accurate measure of effect has been criticised and it is recommended to use more accurate measures such as confidence intervals (Andrade, 2019; Dahiru, 2008; Greenland et al., 2016).

Future research

Further research in this area is warranted. Future studies of this kind may benefit from using vignettes delivered through audio or video. They may also benefit from using information drawn from a real case. Criminal justice workers could also be given a task closer to those experienced in their job role and allocated more time to read the case study through before making decisions. These changes would increase the ecological validity of the study.

Future research should also ensure that the reliability of the questions used is assessed. Additionally, further research is warranted to ascertain whether race-based crime stereotypes held by criminal justice workers in the US and England/Wales are similar, and to explore how these stereotypes might influence their decisions. Furthermore, in terms of assessing for SDR, perhaps the BIDR-16 could be normed on a sample of criminal justice workers before being used, or an alternative SDR measure used.

Finally, it would seem useful to explore whether the race of the criminal justice worker influences their decisions on guilt, blame and punishment. Some mock-juror research has demonstrated ingroup

favouritism, with for example, White mock-jurors showing preferential treatment toward defendants and victims who are of their own racial group (Pfeifer & Ogloff, 2003), whilst other research has found the opposite (Marcus-Newhall et al., 2002), and other no racial bias (Skolnick & Shaw, 1997). It would be interesting to examine whether ingroup-outgroup bias affects the decisions of criminal justice workers. Furthermore, as most criminal justice workers are White, it may be most relevant to explore whether White criminal justice workers' decisions on guilt, blame and punishment are influenced by the race of the defendant and victim. If White criminal justice workers are found to show outgroup bias against BAME defendants particularly, then this may go some way to explaining the disproportionality is the representation of Black people in the criminal justice system.

Conclusion

This study did not find any evidence of racial bias in decisions regarding guilt and blame, in a sample of criminal justice workers practicing in England and Wales. Rather, the perception of guilt and decisions on the type of sentence by CJWs were influenced by their profession. Interestingly, defendant and victim race played a role in the observed profession-related bias in sentencing decisions by CJWs. This may reflect a change in the pattern of racial bias across the CJS in England and Wales over the past decades. Future research should consider the influence of CJW race on decision making. Note, the limitations of this research make it difficult to generalise the results and draw definite conclusions.

Chapter 5: The influence of outgroup bias in the attribution of guilt, blame, and sentencing decisions made by White criminal justice workers in England and Wales.

Introduction

As indicated in previous chapter, no evidence of racial bias in decisions regarding guilt and blame was found in a sample of criminal justice workers practicing in England and Wales. Rather, the attribution of blame and decisions on the type of sentence by criminal justice workers were influenced by their profession. The findings contradict some studies on juror decision making conducted primarily in the US, and specifically using probation officers which found their decisions were influenced by the race of the defendant/probationer (e.g., Bridges & Steen 1998). Such differences might reflect cross-cultural differences between US and England and Wales CJWs in terms of the influence of race on decisions, methodological differences between studies, and or changes in attitudes over time. Alternatively, perhaps the findings from this research indicate that CJWs in the England and Wales CJS are making decisions fairly.

However, questions around disproportionality still arise. The previous study examined the decisions of CJWs collectively. However, simulated studies have found that decisions by mock-jurors can be influenced by ingroup bias, whereby the juror favours the defendant or victim whose racial group matches their own (ForsterLee, et al. 2006; Pfeifer & Ogloff, 2003; Toobin, 1996). Consequently, future research needs to focus on establishing whether criminal justice workers hold similar in-group biases and outgroup/outgroup racial biases to those of other population groups. This is particularly important as when considering the CJS in England and Wales, there is an underrepresentation of BAME people within its workforce compared with their representation

in the general population (Home Office, 2020). If there is racial bias against BAME people in the CJS, given the fact that most legal decisions made about BAME suspects, defendants and victims are made by White CJPs, it is important to assess whether such decisions are fair. If they are found to be racially biased, then this may go some way to explaining the racial disparity in the representation and treatment of BAME within the UK CJS.

Current study rationale and aims.

As most criminal justice workers in England and Wales are White (UK Government, 2021), it is important to ensure that their decisions are not racially biased against BAME suspects, defendants, or victims. Indeed, the influence of race on White criminal justice workers might be different from laypeople, students, and even across criminal justice professions. There is a lack of research in this area using criminal justice professionals as participants and examining whether there are discrepancies in the influence of race between different criminal justice professions. This area of research is important because if the decisions of White criminal justice workers are found to be racially biased, this is likely to have wide implications for the recruitment and training of criminal justice workers.

The above highlight research gaps in evidence (i.e., evidence is contradictory), knowledge (e.g., examining potential bias against Asian offenders and victims or even difference within Asian groups), and population group (i.e., White criminal justice workers and criminal justice professions in the UK).

Although the first empirical study did not find any significant effect of defendant/victim's race on decisions in the entire study population, it did not focus on the specific influence of race among White CJWs. Given the evidence from previous studies on racially driven decision-making of White mock jurors, the second empirical study was conducted to test this possibility. The second empirical study therefore addresses the gap in research by exploring whether the race of the defendant and victim

influenced White criminal justice workers' decisions on guilt, blame and punishment. This is a key difference with the first empirical study. To the author's knowledge this study marks the first of its kind in the field.

Hypotheses:

1. The decisions made by White criminal justice workers regarding guilt, blame and type of sentence will be influenced by defendant's or victim's race.
2. The type of profession will influence White criminal justice workers' decisions on guilt, blame and punishment.

Method

This pre-planned subgroup analysis of White CJWs from study 1 used a quantitative method to assess the influence of victim and defendant race on guilt, blame and punishment decisions in a convenience sample of 399 criminal justice workers in England and Wales. A detailed description of the research methodology has been provided on page 96.

Participants

There were 326 White (criminal justice workers) (163 male and 163 female) with a mean age of 42.3 (*SD* = 11.52) years. Participants were recruited from study one. Participants were mostly working in the police (31.9%) or probation (29.4%) services. However, data on participants' area of practice, that is, racial mix, work location, or sociocultural diversity of the area were not collected.

The study received ethical approval from the University of Nottingham's Faculty of Medicine & Health Sciences Research Ethics Committee (ethics reference No: 289-1803. See Appendix F).

Power

Spencer (2003) investigated racial bias in mock-juror decisions by students and found the effect size of a significant interaction between victim and defendant race to be 0.119 (η^2), a small to medium effect size. Using Cohen's (1988) procedure this converts to an effect size in f^2 of 0.135. Power analysis was conducted for Multivariate Analysis of Variance (MANOVA) in G*Power (Faul et al., 2013) using an α value of 0.05, a β value of 0.80, and an effect size in f^2 of 0.135, indicating a minimum sample size of 88.

Measures

Details of the assessment tools are described in the method study 1, pp 99-100.

Memory Questionnaire (appendix R)

Nineteen non-standardised closed and open-ended questions related to the vignette were used to assess participants' recall of the information in the vignette. Then, participants' responses to these questions were checked against with the nine different scenarios. The answer to the question on defendant race was chosen as the measure of factual memory.

Socially desirable responding

To assess for socially desirable responding amongst participants in the current study, the Balanced Inventory of Socially Desirable Responding-16 (BIDR-16) was used. Furthermore, to help reduce the influence of race saliency and the prospect of socially desirable responding, the Black race-crime congruent crime of assault was used to make race less conspicuous (Sunnafrank & Fonte, 1983).

Variables

Dependent variables

- Blame and guilt (i.e., continuous dependent variables)

- The type of sentence: prison, community sentence, other, and no sentence.

Covariate

- Memory recall of participants was a covariate.

Independent variables

The defendant's and victim's race and criminal justice worker's profession were defined nominally.

Control variable

SDR was measured by total scores on the BIDR-16 questionnaire, ranging from 16 to 112.

The Self-Deceptive Enhancement (SDE) and Impression Management (IM) subscale scores on the BIDR-16 were also defined continuously, ranging from scores of 8 to 56 each.

Procedure

See the method section of study 1, pp 101-102 for details.

The CJW was put in the role of a juror in a case. Defendant and victim race were experimentally manipulated to assess their influence on participants' decisions and memory. Participants read one of nine simulated assault trial transcripts in which the defendant's race (Black/White/ Asian) and victim's race (Black/White/ Asian) were experimentally manipulated (whilst the other facts were held constant) to assess their influence on participants' judgements and decisions.

Statistical analysis

The data was statistically analysed using SPSS version 24.

Data distribution was checked using the Shapiro-Wilk test of normality and visual inspection of histograms. Where distributions were non-normal, non-parametric tests were selected. The α value for significance tests was set at $p < 0.01$.

Since guilt and blame scores were skewed, their correlations with total BIDR score and its subscales were examined using Spearman's test. To test the association between SDR and type of sentence, one-way Analysis of variance (ANOVA) test was used. When necessary, the inferential tests were controlled for SDR.

To measure the main and interaction effects of defendant and victim race as well as participants' memory of case facts on their decisions regarding blame and guilt, a factorial multivariate analysis of variance (MANOVA) was conducted. Correct recalls of the defendant and victim's race were added as covariates to the models. Additionally, significant differences across subgroups of defendant's race, victim's race and criminal justice profession were further analysed using the Bonferroni adjusted method. Partial eta squared was used to measure the effect victim's race, and defendant's race on guilt and blame (Tabachnick & Fidell, 2007). Pairwise comparisons for significant effect were undertaken when required. The influence of defendant's race, victim's race, and criminal justice profession on participants' decisions regarding the type of sentence was tested using multinominal logistic regression.

Results

An overwhelming majority of participants correctly remembered the victim's race (93.8%) and defendant's race (92.8%) per condition. Participants who did not remember the defendant and victim's race was not excluded because only a small fraction of the participants did not remember the defendant or victim's race correctly. However, memory variables were used as covariates in the inferential tests to control for any potential confounding effect of not excluding those participants.

The statistical details of participants' score on measures of SDR on guilt and blame using the BIDR-16 are presented in Table 17 and Table 18.

Table 17

Descriptive characteristics of White CJWs

	Mean	*SD*	Minimum	Maximum
Self-Deceptive Enhancement	37.55	6.49	20.00	55.00
Impression Management	39.00	10.00	15.00	56.00
Total BIDR-16*	77.17	11.78	35.00	109.00
Blame	1.16	.61	1.00	5.00
Guilt	3.83	1.58	1.00	7.00

Note. For all variables *N* = 326

SD: standard deviation

Table 18

Correlation of socially desirable responding with perceived guilt and attributed blame among White CJWs

SDR	Guilt		Blame	
	r	*p*-value	*r*	*p*-value
SDE	.05	.34	.05	.38
IM	-.08	.13	-.05	.40
Total SDR	-.03	.64	-.01	.81

Note. For all variables *N* = 326

SDR: socially desirable responding; SDE: self-deceptive enhancement; IM: impression management

Based on Spearman's correlation test (table 18), decisions on guilt or blame were not significantly related to total SDR, IM or SDE.

Also, SDR total scale, F (3, 322) = 1.06, p = .37, IM, F (3, 322) = .89, p = .37, and SDE, F (3, 322) = .40, p = .75, were not significantly associated with the type of sentence: prison (SDR, *M* [*SD*] = 77.4 [11.3]; IM, *M* [SD] = 39.6 [7.5]; SDE, *M* [SD] = 38.0 [6.0]); community sentence (SDR, *M* [*SD*] = 77.2 [12.1]; IM, *M* [SD] = 40.0 [8.1]; SDE, *M* [SD] = 37.2 [6.5]); other (SDR, *M* [*SD*] = 78.1 [11.3]; IM, *M* [SD] = 40.0 [6.4]; SDE, *M* [SD] = 38.1 [6.9]); no sentence (SDR, *M* [*SD*] = 75.9 [12.9]; IM, *M* [SD] = 39.4 [7.5]; SDE, *M* [SD] = 36.5 [7.0]).

Participant's judgements on guilt and blame by defendant's and victim's race.

As presented in table 19, perceived guilt and attributed blame among White CJWs varied by defendant or victim's race.

Table 19

Perceived guilt and attributed blame among White CJWs by defendant and victim's race

	Guilt		Blame	
Defendant race	Mean	*SD*	Mean	*SD*
Black	3.70	1.65	1.15	.51
White	3.98	1.51	1.10	.45
Asian	3.84	1.57	1.23	.80
Victim's race				
Black	3.87	1.60	1.23	.73
White	3.79	1.65	1.09	.43
Asian	3.85	1.49	1.17	.66
Defendant vs victim's race				
Black - Black	3.60	1.60	1.27	.65

Black - White	3.74	1.71	1.14	.52
Black - Asian	3.80	1.53	1.22	.69
White – Black	4.07	1.45	1.22	.76
White – White	3.79	1.61	1.07	.34
White – Asian	3.91	1.52	1.14	.51
Asian - Black	4.19	1.56	1.19	.74
Asian - White	3.70	1.67	1.13	.62
Asian - Asian	3.64	1.57	1.33	.93

As presented in appendix Q, the defendant's race did not have a significant influence on participants' decisions regarding guilt, F (2, 317) = .88, p = .41) or blame, F (2, 317) = 1.19, p = .030. Likewise, the victim's race was not significantly associated with decisions on guilt, F (2, 317) = .10, p = .90) or blame, F (2, 317) = 1.58, p = .21). Also, the defendant's and victim's race did not show significant interactions on decisions regarding guilt, F (4, 317) = .49, p = .74 or blame, F (4, 317) = .31, p = .87.

Participants' judgements on the suggested sentence by defendant's and victim's race

Multinominal logistic regression models were created to evaluate the association of the defendant's and victim's race with the suggested sentence as an outcome variable with more than two levels. In the regression models, the defendant and victim's race were included separately, with White being set as the reference category. Also, no punishment was defined as the reference category for the outcome. According to the multinominal logistic regression models, there was no significant association between the defendant's, X^2(6, N = 326) = 4.61, p = .59, or victim's race, X^2(6, N = 326) = 13.32, p = .03, with the sentencing decisions.

As shown in table 20, describing the defendant as Black versus White did not change the likelihood of suggesting a community or prison sentence among White CJWs. Likewise, the likelihood of suggesting

a community, or prison sentence, was comparable for a defendant described as Asian vs White among White CJWs. Similarly, describing the victim as Black versus White did not change the likelihood of suggesting a community, or prison sentence among White CJWs. Also, the likelihood of suggesting a community, or prison sentence was comparable for a victim described as Asian vs White among White CJWs.

Table 20

Sentencing decisions by White CJWs across defendant and victim race.

	Community Sentence			Prison			*Other Sentence*		
	OR	95%CI [LL, UL]	*p*	OR	95%CI [LL, UL]	*p*	OR	95%CI [LL, UL]	*p*
Defendant's race									
Black	.75	[.34, 1.67]	.49	.89	[.437, 1.83]	.76	.75	[.33, 1.71]	.49
Asian	1.01	[.43, 2.33]	.98	1.60	[.76, 3.33]	.22	.89	[.37, 2.15]	.796
White (reference)									
Victim's race									
Black	1.91	[.86, 4.23]	.11	1.54	[.76, 3.09]	.23	.49	[.20, 1.21]	.12
Asian	1.75	[.75, 4.08]	.19	1.97	[.95, 4.05]	.07	1.19	[.52, 2.70]	.676
White (reference)									

Note. The reference sentence category is no punishment.

White CJWs' decisions on guilt and blame by criminal justice profession.

As presented in Table 21, perceived guilt and attributed blame of White CJWs varied across professions.

Table 21.

Perceived guilt and attributed blame by White CJWs' professions

Criminal justice professions	*n*	Guilt		Blame	
		Mean	*SD*	*Mean*	*SD*
Other criminal justice professionals	14	3.79	1.63	1.14	.36
Police Worker	104	3.86	1.86	1.10	.41
Prison Worker	42	3.74	1.48	1.29	.77
Probation Worker	96	4.21	1.27	1.19	.73
Youth Justice /Social Work	70	3.36	1.47	1.14	.62

SD: standard deviation

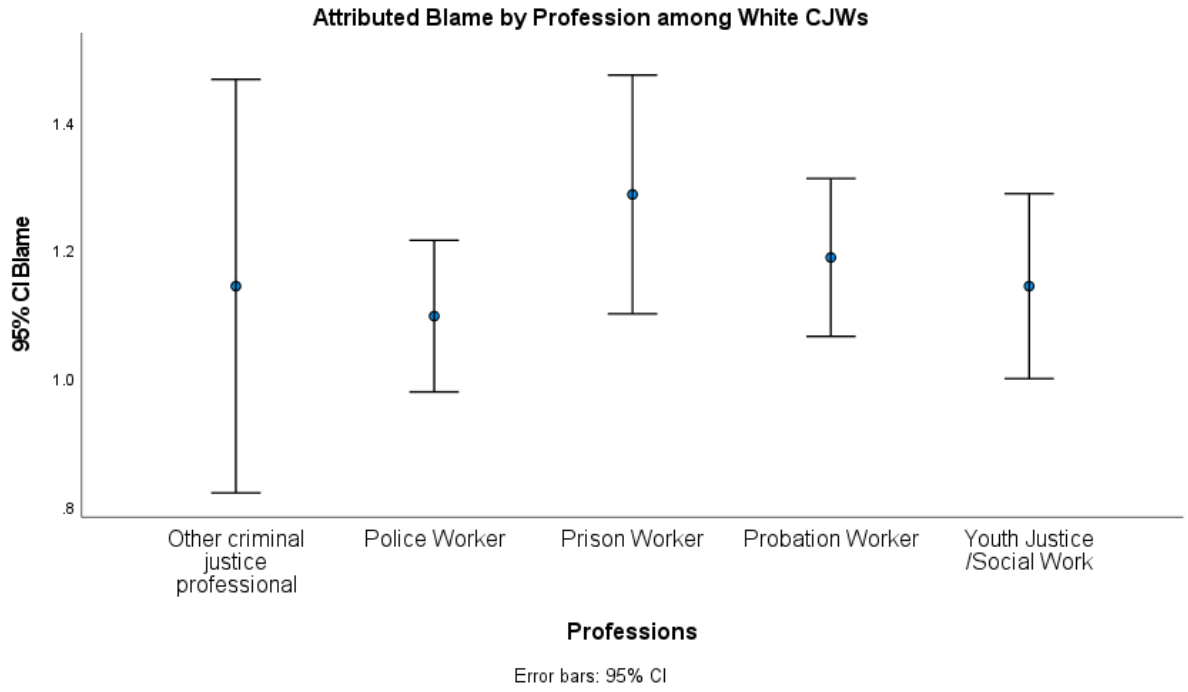
The effect of CJWs' profession on perceived guilt and attributed blame was analysed using MANOVA (Appendix S).

Based on MANOVA results, Pillai's trace F (8, 642) = 1.94, p = .05, profession did not significantly influence White CJWs' decisions on blame, F (4, 321) = 1.60, p = .17. In contrast, participants' decisions on guilt were significantly affected by their profession, F (4, 321) = 3.79, p = .005. After the Bonferroni

corrected post-hoc analysis, probation workers and youth justice/social workers were found to make significantly different decisions on guilt (*MD* = 0.85, *p* = .006). Thus, probation workers attributed a higher degree of guilt to the defendant as compared to youth justice or social workers (figure 5).

Figure 5

Variation in the attribution of blame across criminal justice professions among White CJWs



Participants' decisions on the type of sentence by criminal justice profession

The proportion of CJWs who chose a type of punishment (community sentence, prison, or another sentence) rather than no punishment varied across the CJWs' professions (table 22).

Table 22.

Suggested sentence by profession among White CJWs

Professions	**Type of Sentence**			
	No sentence	Prison	Community Sentence	Other
Other (%)	33.3%	25.6%	23.1%	17.9%
Police (%)	22.1%	51.0%	5.8%	21.2%
Prison (%)	21.4%	42.9%	19.0%	16.7%
Probation (%)	11.5%	42.7%	28.1%	17.7%
Youth Justice /Social Work (%)	31.1%	15.6%	42.2%	11.1%

Further, multinominal regression models showed a significant association between criminal justice profession and sentences suggested by White CJWs, $X^2(12, N = 326) = 58.29$, $p < .001$. As compared to youth justice/social workers, police (OR = 5.03; 95% CI = 2.12, 11.94; $p < .001$), probation (OR = 8.13; 95% CI = 1.50, 12.75; $p <.001$), and prison workers (OR = 4.36; 95% CI = 3.06, 21.58; $p = .007$) were more likely to choose prison sentence over no punishment. Conversely, police workers were less likely to choose community sentences over no punishment compared with their colleagues in youth justice or social services (OR = .23; 95% CI = 08, .66; $p = .007$).

Table 23

Sentencing decisions by White CJWs across criminal justice professions

	Community Sentence			Prison			*Other Sentence*		
Profession	OR	95%CI [LL, UL]	*p*	OR	95%CI [LL, UL]	*p*	OR	95%CI [LL, UL]	*p*
Other	.30	[.03, 3.04]	.31	4.36	[.92, 20.74]	.064	4.00	[.73, 21.83]	.007
Police	.23	[.08, .66]	.007	5.03	[2.12, 11.94]	<.001	2.87	[1.06, 7.73]	.11
Prison	.79	[.26, 2.37]	.67	4.36	[1.50, 12.75]	.007	2.33	[.65, 8.32]	.04
Probation	2.18	[.90, 5.32]	.09	8.13	[3.06, 21.58]	<.001	4.63	[1.54, 13.96]	.19
youth justice/ social workers (reference)									

Note. The reference sentence category is no punishment.

CI: confidence interval; OR: odds ratio

Discussion

This study explored the possibility of racial bias amongst White criminal justice workers in England and Wales using a mock-jury paradigm.

Perceived guilt, attributed blame, and suggested sentence by the race of the defendant and victim

This study did not find any significant association between the defendant's race and perceived guilt, blame attribution, and sentencing decisions by White CJWs. This finding has mixed precedent. The result fits with that of Skolnick and Shaw (1997) and the England and Wales simulated study by Thomas (2010), which both found no significant difference in White mock jurors' treatment of Black compared to White defendants. However, the non-significant finding is inconsistent with some studies conducted outside of England and Wales. For example, the meta-analyses by Mitchell et al. (2005) and Devine and Caughlin (2014), found that mock-jurors tended to show more leniency toward defendants of their racial ingroup. The finding is also inconsistent with the Canadian study by Pfeifer and Ogloff (2003), which found favourable ingroup bias toward both White defendants and victims among White mock-jurors. It also contrasts with that of Marcus-Newhall et al. (2002), who found favourable outgroup bias toward Black compared to White defendants among White mock jurors. The difference in findings might reflect cross-cultural differences in terms of the influence of race on decisions.

The present study also did not find a significant link between the victim's race and perceived guilt, attributed blame, or suggested sentence among White CJWs. This conflicts with the reported association between victim race and the bias in mock-juror decisions (e.g., FosterLee et al. 2006). In a study of White Australian mock-jurors, FosterLee et al. (2006) found favourable outgroup bias when the victim was known to be Black. Interestingly, a UK study found that White mock-jurors from more racially diverse neighbourhoods were more likely to show ingroup bias against White defendants of interracial crimes (Thomas, 2010). The latter can be explained via the 'Black sheep effect' and is said

to occur as members of the ingroup want to distance themselves from the behaviour of the deviant ingroup member (Marques et al, 1988). However, since the information about the diversity of the practice area was not included in the present research, it is difficult to make inferences about the contribution of diversity to the observed lack of ingroup bias in criminal justice workers.

Participants' judgements on blame, guilt, and punishment by profession

In the current study, CJWs' profession significantly influenced their decisions on the degree of guilt, with probation workers suggesting a higher degree of guilt to the defendant as compared to youth justice or social workers. However, blame attribution was not related to CJWs' profession. In a US study (Bridges & Steen, 1998), probation officers tended to attribute crimes committed by Black defendants to internal causes whilst crimes committed by White defendants were mostly attributed to external factors. This profession-related variation in decisions on guilt vs blame may indicate cultural differences in the attribution of blame between probation officers from England and Wales and the US. Further, the researcher used content analysis to explore racial bias in probation officers, whereas the present study relied on case vignettes and recruited a broader scope of criminal justice professions.

Additionally, the present study found that suggested sentences were influenced by CJWs' professions. Among White criminal professionals, youth justice/social workers were the most lenient group regarding sentencing decisions. In comparison, probation workers police officers, and prison workers were more likely to suggest a harsher punishment for a Black-race stereotypical crime. This is consistent with Graham and Lowery's (2004) US study which compared the decisions of 105 ethnically diverse police officers and 91 juvenile probation officers on blame, punishment, and recidivism. They found that police and probation officers primed with Black race stereotypical words were more likely to view defendants (with unspecified race) as blameworthy, recommend severer punishment, and predict an increased likelihood of recidivism, than non-primed officers. Likewise, Gordon (1993) found

that when a crime was race congruent, mock-jurors were more likely to perceive the defendant as guilty and recommend severer punishments than when the crime was race-incongruent. Furthermore, in the Sunnafrank and Fontes (1983) study, undergraduates were used as participants rather than CJWs. It is possible that CJWs do not hold race-based crime stereotypes; or if they do, their stereotypes are different from undergraduates. Therefore, the findings in the current study might be because the crimes used are not race-stereotypical crimes for CJWs. However, the observed associations between profession and sentencing decision were not adjusted for by defendant's or victim's race. Additionally, the higher odds of choosing severe punishments in criminal justice workers from other disciplines might merely reflect the small number of participants from other professions.

Possible explanations for the inconsistency in the current and previous research findings.

Differences in racial bias. An explanation for the non-significant result may reflect a decline in racial bias over time. Perhaps people and society are less biased than when previous research was conducted, or professional standards, training and processes have improved. Perhaps CJPs in England and Wales are not racially biased. Alternatively, it is possible that individuals might have become more sophisticated or covert in their racist attitudes and beliefs, which itself might be influenced by laws and policies directed at curtailing biased decisions and a personal desire to appear unbiased (Dovidio & Gaertner, 2004). Also, criminal justice workers might be differently biased than the general population suggesting that these CJWs may have less racial discrimination. This highlights the need for a comparison group, preferably with members of the public or with a jury eligible population group. Similarly, an explanation for the discrepancies in findings might be due to cultural differences, with many simulated studies having been conducted in the US. It is possible that the differences reflect real cultural differences in racial bias between England and Wales and US mock-jurors, with an England and Wales sample population being less racially biased. The role of extra-legal factors should also be

explored because the interconnection of these factors with race in influencing decisions by criminal justice workers has not been specifically studied. The latter perhaps highlight that it might also be of interest to compare how large an effect racism has on criminal justice workers in comparison to other sources of bias, like gender, sexuality, nationality etc.

The strength of race saliency. Race saliency has been observed to influence the expression of racial bias (Sommers & Ellsworth, 2001). Specifically, Sommers and Ellsworth (2001) found that when the race of the defendant was made salient, White mock-jurors were less likely to be racially biased. The race of the defendant and victim was mentioned several times throughout the vignette, the obviousness of which may have prompted participants to supress any potential racial bias (cf. Dovidio & Gaertner, 2004). A more egalitarian value system may have developed in many Western societies, which has prompted people to want to be or appear less biased (Axt et al., 2016). Perhaps then, the present study's non-significant result was influenced by the amount of race saliency in the vignette. If race was made less salient different results might have occurred. Although the study used the BIDR-16 to check for social desirability in participants, it could have also tried to address it through using the perspective-taking method: asking participants to provide opinions based on what they think the average person would do, rather than what they would do themselves. In this instance it would be asking participants to rate what they thought the average CJP would do if presented with the case. Past research has shown that using a perspective taking method reduces social desirability biases (Cox & Devine, 2015; Calabrese et al., 2017). The idea is that participants project their prejudices onto the 'average person', without having to personally admit to holding these views themselves (see Krueger & Clement, 1994).

Additionally, race could also have been made less salient through reducing the number of times that the race of the defendant and victim is mentioned in the text, and instead using pixelated photographs of the defendant and victim, so that only their racial identity could still be seen by the study participant. Some past research has found that racial stereotypes are more likely to be activated by more vivid information (e.g., photographs), than written/verbal categories or labels (Beckett & Park, 1995). However, making race less salient would perhaps reduce the ecological validity of the study, as decisions made by CJWs are generally made with the knowledge of the race of the defendant (and victim, if there is one).

The strength of the evidence. The strength of the evidence seems to affect legal decisions (Charman et al., 2017; Whitestone et al., 1995). Indeed, fundamentally legal decisions should be based on the strength of the evidence rather than extra-legal factors. Consistent with the *liberation hypothesis* (Bjerregaard et al., 2017; Kalven et al., 1966), research by Gordon (1993) found that the race of the defendant tended to be a factor in mock-juror decisions when the evidence against the defendant was weak, but in circumstances when the evidence against the defendant was strong decisions on guilt and punishment were based largely on the evidential facts. Consequently, it is possible that the present study's non-significant result is the consequence of the evidence against the defendant being too compelling. However, Gordon (1993) indicated that race-crime congruency overrode the influence of strength of the evidence.

Differences in crime-race congruences/ stereotypes. Another possible reason for the discrepancy in results relates to the selection of the race-stereotypical crime. The current study used the crime of assault to explore racial bias in the decisions of criminal justice workers, based on a US based study by Sunnafrank and Fontes (1983). They used a population of 78 undergraduate students who were required to match photographs of five Black male and five White male defendants to a list

of 10 crimes. They found that Black males were stereotypically associated with the crime of assault. It is worth noting that Sunnafrank and Fontes did not identify the ethnicity/racial group or gender of participants used in their study. If the participants used were of only one given group, then this limits the generalisability of the findings. It is possible that race-stereotypical crimes might not exist within certain groups or differ across populations; and/or might have changed within the population group used in the Sunnafrank and Fontes study. For example, in 1970s and 1980s Britain, a stereotype for a terrorist was perhaps likely to be someone of Northern Irish origin and supporter of the Irish Republic Army (IRA) or; whilst following events including the London bombings in 2005 and in 2017 the Manchester bombing, and the London Bridge and Borough Market terrorist attacks, stereotypes around the profile of terrorists is likely to have shifted to someone who is an Islamic extremist trained by Islamic State or al-Qaeda (cf. Ferreira, 2021). It is worth noting that unlike Sunnafrank and Fontes (1983), Thomas (2010) did not find the crime of assault to be a race stereotypical crime for Black defendants amongst a sample of jury eligible participants recruited in England and Wales. Perhaps then, the crime selected was not race-stereotypical enough to subliminally influence the decisions of participants. Alternatively, perhaps using a crime stereotypical for Black offenders helped raised race saliency, which as outlined earlier has been found to reduce the likelihood of racially biased responses (Sommers & Ellsworth, 2001). Research is needed that compares the power of race saliency against race-crime congruency, particularly in the context of criminal justice professional decisions.

Methodological issues

This study has the same methodological issues as study 1 (see study 1 pp. 120–125), and some more of its own. For example, the use of a comparator group of CJWs from minority ethnicities could have improved the understanding of the race-influenced decision-making by White criminal justice workers in this study, particularly for in-group bias and race-crime congruency. In terms of the vignette, using a combination of race-crime congruent and incongruent conditions and gender would have been more illuminating. The race saliency of the scenarios is another extra-legal factor which may change decision-making by participants. However, it might be difficult to fit all these variables into one experiment.

Another methodological issue with the current study is that it did not assess whether the crimes selected were race stereotypical for CJWs in England and Wales. As stated earlier, it is possible that CJWs in England and Wales hold different or no crime-based race stereotypes. It perhaps would have been better if the current study first had a sub-study that replicated the Sunnafrank and Fontes study using CJWs as participants. If a stereotypical crime were identified, then it could have been used to assess its influence on White CJWs' decisions.

According to the United Nations there are 48 countries in Asia. The term 'Asian' therefore covers a wide and heterogenous category of people, and the use of the term might conjure up different ethnic groups and stereotypes for participants. The diversity in categorisation of 'Asian' might have impacted on the CJWs' judgements in this study. Interpreting a defendant/ victim as 'South Asian' (e.g., Pakistani), might result in different judgements compared to interpreting a defendant/ victim as 'East Asian (e.g., Japanese). Consequently, it might have been better if this study was more specific in its use of the term 'Asian'. However, the same might be said of the terms 'Black' and 'White'.

Finally, a possible methodological issue with the current study has been its exclusive use of a quantitative method. Although quantitative methods have several advantages, adopting a mixed methods approach where qualitative research methods are also employed, might have added more depth, and understanding to the subject area. If racism is made subtler, then using interviews could have been used to explore further this subtleness, including CJWs' perceptions on racism; possible levels of racism and practices that they have observed across their working life, and how this might have changed.

Future research

It is strongly recommended that the possibility of in-group/outgroup bias in non-White criminal justice workers are compared to White criminal justice workers to detect the ongoing trends more clearly. Furthermore, it would be useful to assess as a sub-study whether White criminal justice workers hold specific race-crime stereotypes, and then use that crime to check whether crime-race congruency does influence their decisions – or simply ingroup/outgroup bias. The same research can also be applied to non-White CJWs. Additionally, a wider scope of crimes should be explored to identify the variation in racial bias across offence types and defendant-victim race combinations. The latter would also require more thinking around the types of questions used to study and assess whether racial bias impacts on the processing, (mis)remembering and ultimately influences the decision making of participants. It might also be interesting to explore how other identified extra-legal factors interact with race to influence decisions (e.g., the juror's/assessor's/ defendant's/ victim's gender, sexual orientation, religious belief). Similarly, future research could explore how the strength of the evidence interacts with race to influence the decisions of CJWs. Also, future research could provide CJWs with tasks that are more relevant to their actual work, which might provide a greater insight into their decision and allow for the results to be more easily generalised.

Additionally, as is the direction of some research in North America, it might be worth examining White CJWs' use of risk assessment tools (e.g., Structured Assessment of Violence Risk in Youth [SAVRY], Borum et al., 2006), to assess whether such instruments (commonly used by CJWs in England and Wales) are racially biased.

Furthermore, to restate an earlier comment, many of the studies within the field were conducted several decades ago. Therefore, it might be useful if future research examines the publication dates of these studies conducted, to assess whether racial bias has changed over time. This stance would review whether there have been any cultural changes.

Finally, it is suggested that future research use a mixed method approach as this might provide a richer understanding of the influence of race within the CJS.

Conclusion

The present study suggests that White criminal justice workers' decisions on guilt, blame, or type of sentence are not influenced by defendant or victim race. In contrast, perceived guilt, and suggested sentence by CJWs appear to be partly driven by their professions. Therefore, this research does not support the possibility of ingroup or outgroup bias in perceived level of guilt and blame among White criminal justice workers in the context of the Black-race stereotypical crime assault. Further, it seems unlikely that sentencing decisions by this sample of White CJWs is influenced by defendant race. However, White CJWs' profession appears to influence their suggested sentence, with probation and police officers tending to be harsher in their decisions than youth justice/social workers. This has an important implication for the assessment of racial bias among White CJWs in different areas of practice. In addition, the study found some evidence, though non-significant, for the influence of victim's race on sentencing decisions by White criminal justice workers, with severer types of sentences being considered for defendants whose victims were described as Asian. These points are

important when considering CJWs' contribution to the final decision regarding imprisonment or release. The findings from this research suggests that differences between races regarding arrest and imprisonment might not be about CJS bias, but rather the proportion of crimes committed by the various BAME groups. This requires further research into the drivers of disproportional crime rate among BAME groups.

Chapter 6: Discussion, conclusions, and future research

This thesis addressed the gap in research and knowledge involving the influence of race of defendant and victim on CJWs' decisions on guilt, punishment/sentencing, and blame. It used a simulated juror method to explore the possibility that race of the defendant, the race of the victim, the race of the criminal justice worker (CJW), and the criminal justice profession might play a role in the decisions of CJWs. The thesis started with a systematic review of the literature between January 2015 and November 2021 that explored the influence of Black versus White defendant and victim race on mock juror perceived guilt, sentencing/punishment decisions, and blame attribution using the simulated mock-juror paradigm. The systematic review concentrated on decisions between Black versus White defendants and victims as a scoping exercise of the literature in the area prior to setting the review question, identified a lack of research relating to decisions about Asian defendants and victims. The lack of research about decisions regarding Asian defendants and victims might reflect the fact that Asian people are often underrepresented in the CJS compared with the representation in the general population. Still, research in this area is important to undertake as Black people are disproportionately represented in the England and Wales CJS. Racial bias might explain some of the racial disproportionality figures.

The systematic review highlighted contradictory finding on whether mock jurors show racial bias in their decisions concerning Black versus White defendants and victims, as well as highlighted a need for research that considers views about Asian defendants and victims. Furthermore, the systematic review identified the possibility of risk of bias. Risk of bias is particularly important when administering self-report questionnaires, particularly in relation to socially desirable responding whereby participants might provide answers that present themselves in a positive light. Such socially desirable responses would invalidate the research findings. As this thesis involved getting criminal justice workers to complete self-report questionnaires it was vital to have a measure in place that could assess socially

desirable responding amongst the participants used in the current research. Consequently, chapter two critically reviewed the BIDR-16 psychometric developed by Hart et al. (2015) to detect and control for socially desirable responding amongst participants as a source of bias in population studies. Overall, a review of the psychometric properties of the original 40-item BIDR has shown it to have satisfactory internal consistency, temporal stability, and discriminant and convergent validity (Paulhus, 1994). However, there is some overlap of SDE and IM with external correlates, which limits the external validity of these BIDR subscales. Although the overall practicality of the BIDR-16 makes it an attractive substitute for the longer BIDR, the construct validity and internal consistency of the items are not optimal, and the BIDR-16 should be used with caution until further refinement and supplementary studies. Consequently, the BIDR-16 was therefore used in the current thesis with caution to assess for socially desirable responding amongst participants/ CJWs.

The thesis then went on to examine the role of race of defendant, victim, and criminal justice worker on the decisions of CJWs. It seems important to study the decisions of CJWs as they are involved in all stages of the criminal justice process. The matter is particularly exemplified for White criminal justice workers, who are particularly influential in the decision-making process considering that they constitute most of the CJWs in England and Wales. As socially desirable responding (SDR) is an issue when administering self-report questionnaires, as SDR could invalidate the results, the BIDR-16 was used to detect the possibility of SDR amongst participants. The overall findings from this thesis are discussed below.

The systematic review of this doctoral research did not find compelling evidence for the differential treatment of Black versus White defendants or victims by mock-jurors in terms of guilt, sentencing, and blame. Some studies reported that White defendants were most likely to be found guilty, while a few studies demonstrated that Black defendants received more lenient sentences compared to White

defendants. Two out of three studies found victim race did not significantly affect mock jurors' decisions of blame for a sexual assault offence. In addition, some studies showed that mock jurors were more likely to blame White than Black defendants for an offence. The systematic review highlighted the gap in the literature, as there is limited research examining the influence of defendant and victim race on CJWs' decisions of guilt, punishment/sentencing, and blame. In the context of previous research, these findings are consistent with the UK study by Thomas (2010) and a meta-analysis by Mazella and Feingold (1994), which found no evidence of racial bias in sentencing decisions. Similarly, the finding that there was no clear evidence of discrimination against Black versus White defendants on measures of guilt is consistent with the findings of some previous meta-analyses (Devine & Caughlin, 2014; Mazella & Feingold, 1994; Mitchell et al., 2005). However, the finding of no significant interaction between mock juror race and defendant race in this systematic review is inconsistent with the meta-analysis of Devine and Caughlin (2014), which found Black mock jurors showed outgroup bias in their guilt decisions against White defendants, whilst White mock jurors showed no significant racial differences in their guilt decisions. As stated in the systematic review chapter, the findings from the current systematic review might indicate a decline in racial bias over time or reflect slight differences in the inclusion/exclusion criteria between the present systematic review and previous meta-analyses.

This thesis found mixed evidence regarding the influence of defendant and victim's race on decisions of guilt, sentencing/punishment and blame by CJWs.

The first empirical study in this thesis found that the defendant and victim race did not significantly affect the decisions of criminal justice workers on guilt, blame and sentencing. However, perceived guilt and suggested sentence were significantly influenced by criminal justice professions. Probation workers suggested a significantly higher degree of guilt to the defendant compared with youth

justice/social workers. Youth justice/social workers were also least likely to suggest a prison sentence as compared to other professionals. In addition, there was some evidence of an interaction between race and criminal justice professions. In this study, the effect of profession on the type of sentence was modified by defendant and victim's race: when the victim was described as Black or Asian, youth justice workers/social workers were more likely to suggest community sentence compared to other professionals. Contrastingly, police workers were most likely to suggest a prison sentence when the victim was described as White, whilst youth justice /social workers were least likely to suggest a prison sentence when the victim was described as Black. Furthermore, decisions on blame, and type of sentence were directly influenced by criminal justice professions. This is a new area of research and warrants further research to bring further clarity and understanding of the role of profession on decisions. It is also important to explore the influence of the race of the CJW on their decisions, particularly as previous mock juror research has highlighted the possibility of ingroup and outgroup racial bias in decision making (e.g., Toobin, 1996).

Following on from the first empirical study, the second empirical study explored the issue of ingroup and outgroup racial bias in the attribution of guilt, and blame, as well as sentencing decisions made by CJWs. Overall, this study did not find evidence of racial bias by White criminal justice workers on decisions of guilt, sentence, and blame. The result fits with that of Skolnick and Shaw (1997) and the UK simulated study by Thomas (2010), which both found no significant difference in White mock jurors' treatment of Black compared to White defendants. Interestingly, the other studies did not use CJWs as participants, but the similar non-significant results suggest that CJWs' views might not be significantly different from the views of non-criminal justice workers. However, both the Skolnick and Shaw (1997) and Thomas (2010) studies were conducted some time ago, and more research using both

CJWs alongside non-CJWs as participants needs conducted to confidently ascertain whether there are significant differences in verdict, sentencing and blame attribution across these different groups.

Note however, the non-significant finding in study two is inconsistent with the meta-analyses of Mitchell et al. (2005) and Devine and Caughlin (2014), who found that mock-jurors tended to be more lenient towards defendants of their own racial in-group; and is also inconsistent with the findings of FosterLee et al. (2006) who found White mock jurors showed favourable outgroup bias when the victim was known to be Black. The reason for such inconsistency in findings might be because of cultural differences, as the current study used participants recruited in the England and Wales and the other studies did not. Alternatively, it could be because of differences in profession, as the current study used CJWs, whist the other studies did not.

The second empirical study also found some evidence on the influence of criminal justice professions on perceived guilt and suggested sentence by White CJWs. Note however, the race of defendant and victim did not have a main effect on the outcome variables but interacted with profession on the suggested sentence. Specifically, probation workers tended to perceive a significantly higher degree of guilt for the defendant as compared to youth justice/social workers. Also, probation workers and police officers were more likely to consider prison sentence for the defendant compared to youth justice/social workers. This is new area of research, and more studies need to be done to shed further light and understanding on the decisions of CJWs. For example, new research could examine the influence of CJWs' race within each profession on their decisions, for example, are there differences between Black versus White police officers' decisions, compared to Black versus White probation officers.

The theoretical implications of thesis research findings.

Social Identity Theory (SIT).

There are several implications of this thesis findings to theories on racial bias, including Social Identity Theory (SIT). As stated in chapter five, key elements of the SIT are the concepts of ingroup and outgroup bias (Tajfel & Turner, 1986). To recap, according to SIT the group that someone identifies with is termed 'ingroup', and all other groups that fall outside that ingroup is termed 'outgroup'. Consequently, the world can be viewed as divided into them and us. SIT stipulates that people naturally try to present a positive image of their ingroup and uses the term 'positive distinctiveness' to describe the process of people trying to create a positive self-image of their ingroup in comparison to the outgroup. So, applying SIT to the context of possible juror racial bias, SIT predicts that mock jurors would show racial bias or favouritism towards defendants/victims who share their racial ingroup over defendants/victims who are from a racial outgroup.

Contrastingly, a derivative of SIT is the 'black sheep' effect, which is when someone treats members of their ingroup less favourably than those from the outgroup (Marques et al., 1988). This is said to occur to distance the actions of that ingroup member from the ingroup, so that the overall positive image of the ingroup is preserved (Marques et al., 1988). The deviant ingroup member is rendered an outcast. Consequently, the 'black sheep' effect has been used to explain why some (mock) jurors show outgroup favouritism in decisions on guilt, blame and punishment.

Overall, the non-significant finding in study two that White CJWs' decisions on guilt, blame and punishment were not affected by the race of the defendant nor the victim, is inconsistent with the SIT position on ingroup-outgroup bias and the 'black sheep' effect. Nevertheless, it is worth noting that study two found that criminal justice worker profession played a role in decisions on guilt rather than

the race of the defendant or victim. Specifically, probation workers tended to perceive a significantly higher degree of guilt for the defendant as comparted to youth justice/social workers. This finding contrasts with study one, which found criminal justice worker profession interacted with defendant and victim race to impact on sentencing decisions (e.g., police workers most likely to suggest prison when the victim was described as white). Perhaps the BAME CJWs' decisions affected the overall findings in study one. The contrasting finding highlights the need to research the views of BAME CJWs as it is possible that BAME CJWs do social racial bias.

Race stereotypical crimes.

As stated in the previous two chapters, the crime of assault was identified by Sunnafrank and Fontes (1983) as a race stereotypical crime for black defendants and predicted that defendants accused of a racial stereotypical crime was more likely to be found guilty and perceived as more blameworthy for the offence occurring than defendants whose race is not stereotypical for the offence. Regarding blame, some research has supported the assertions that attributions regarding the cause of behaviour are partly influenced by having crime stereotypes (Bodenhausen & Wyer, 1985); race stereotypical-consistent behaviour being more likely perceived dispositional (Taylor & Jaggi, 1974); and more likely to be generalised to other situations and across time than is non-stereotypical consistent behaviour (cf. Bodenhausen & Wyer, 1985). Furthermore, there is evidence that transgressions attributed to stable causes are more likely to receive harsher punishments than those viewed as caused by transitory factors (Bodenhausen & Wyer, 1985). Indeed, research by Gordon (1990) found that a black defendant accused of the race stereotypical crime burglary, was more likely to be found guilty and blamed for the offence occurring than their victim, compared with when they were accused of committing a non-stereotypical crime. Note, in this thesis the crime the defendant was accused of committing was assault, which has been identified as a crime stereotypically committed by Black defendants

(Sunnafrank & Fontes, 1983). Based on previous research, it would be predicted that a Black defendant accused of the Black stereotypical crime assault is more likely to be found guilty and perceived as more blameworthy for the offence than a white defendant accused of assault. Neither study one or two found CJWs' decisions on guilt, blame and sentencing were affected by the race of the defendant or victim. These non-significant findings might reject the concept of race stereotypical crimes or be the result of cultural differences in what crimes are race stereotypical in the US and England and Wales. Perhaps assault is not a race stereotypical crime for Black people in England and Wales? However, as criminal justice profession significantly influenced decisions on guilt and suggested sentence, perhaps certain professions are more influenced in their decisions on guilt and sentencing by the race stereotypical crime of assault than other professions. Future research should consider whether in England and Wales racial stereotypes on who commits certain crimes should be explored; whether criminal justice profession hold crime stereotypes based on race; and whether if crime stereotypes do exist amongst CJWs, the influence of these stereotypes vary across criminal justice professions. Furthermore, as this thesis did not explore the decisions of just Black CJWs, it would be useful to explore whether Black CJWs hold stereotypes about which racial group commit crimes, and whether this influences their decisions on guilt, blame and punishment. If some BAME CJWs are behaving discriminatorily towards BAME defendants and victims who share their own racial ingroup, then some of the reason might be pressure to avoid being seen as lenient towards people of their own racial ingroup. Such behaviour could be considered a similar to the black-sheep effect.

Methodological issues.

There are some concerns raised in the recruitment procedures, participants, methodology and statistical procedures employed in the various studies within this thesis. Collectively, the above methodological issues ultimately impact on the overall generalisability of these studies.

A possible methodological issue of the systematic review is its use of the CEBMa critical appraisal tool, which has not been used in research as much as the CASP and AXIS critical appraisal tools. The lack of use of the CEBMa tool in research could be considered a limitation. Nevertheless, the CEBMa does cover many of the domains used in the CASP and AXIS and can be considered a compromise between the shorter CASP tool and longer AXIS tool. Furthermore, at the time of this systematic review, there were no known validated critical appraisal tools for cross sectional studies.

Methodological concerns for both empirical studies includes the use of an online survey, which might have increased the possibility of selection bias through obtaining an overrepresentation of participants who were educated, White and female (Benedict et al., 2019; Whitaker, 2017); and the use of the mock juror paradigm itself, with the artificial use of case vignettes raising concerns around ecological validity as vignettes might disconnect participants from the case (Bornstein et al., 2017; Paddam, et al., 2010). Regarding the last point, the fact that participants were aware their decision had no real-life consequence might have affected their decisions and have less meaning compared to if their decisions had real life consequences (Bornstein et al., 2017). Additionally, the lack of qualitative element to the empirical studies could be considered a limitation. Using a mixed methods approach that incorporated qualitative methods, might have added more depth, and understanding to the subject area. If for example, racism is more subtle, then using interviews could have been used to explore further this subtleness, including CJWs' perceptions on racism; possible levels of racism and practices that they have observed across their working life and how this might have changed.

Importantly also, there are some statistical issues to consider including: the BIDR-16 measure has not been validated on a criminal justice population, which questions the appropriateness of its use in studies one and two; and there was not a reliability analysis of the survey questionnaires used to measure CJWs' judgements on guilt, blame and punishment.

Recommendations

Based on the findings and methodological limitations of this doctoral research, the following recommendations are made:

- Future studies should incorporate a qualitative element to explore further CJWs' perceptions of racism; and possible levels of racism and practices that they have observed across their working life and how this might have changed.
- To increase ecological validity, vignettes should use information drawn from real cases and be delivered through audio and/or video.
- CJWs should be provided with tasks that are more relevant to their actual work, which might increase ecological validity of the study, provide a greater insight into their decision, and allow for the results to be more easily generalised.
- The reliability of questions used in the studies should be assessed.
- Further research is warranted to ascertain whether race-based crime stereotypes held by CJWs in the US and England/Wales are similar, and to explore how these stereotypes might influence their decisions.
- Research samples should have a more equal gender and race balance across experimental conditions and professions – and include in their analysis responses by gender and race, so to assess whether these variables effect decisions.
- To assess for SDR, perhaps the BIDR-16 could be normed on a sample of CJWs before being used, or an alternative SDR measure used.
- The possibility of in-group/outgroup bias could be studied with a sample of non-White criminal justice workers and compared to White criminal justice workers to detect ongoing trends more clearly.

- A wider scope of crimes used to explore and identify the variation in racial bias across offence types and defendant-victim race combinations.
- As is the direction of some research in North America, it might be worth examining White CJWs' use of risk assessment tools (e.g., Structured Assessment of Violence Risk in Youth [SAVRY], Borum et al., 2006), to assess whether such instruments (commonly used by CJWs in England and Wales) are racially biased.

Conclusion

In conclusion the research in this thesis found inconsistent evidence for differential treatment of defendants and victims on decisions of guilt, sentencing/punishment and blame based on their race. Some of the research findings in this thesis contradict racial disproportionality figures that indicate an over-representation of certain BAME groups in the CJS compared to their representation in the general population and compared to their White counterparts. Consequently, factors other than racial bias by jurors might account for racial disproportionality in the CJS. First, the disproportionate figures might reflect the disproportionate amount of actual crime committed by BAME versus White defendants. Second, the racial discrepancies in sentence length and type might be the consequences of Black defendants being more likely to plead not guilty compared with White defendants. Note, judges can take up to one third off sentences for defendants who plead guilty. Therefore, if Black defendants plead not guilty but are found guilty, judges cannot reduce their sentences. Consequently, the disproportionality in the sentencing of Black defendants might reflect Black defendants' decisions to plead guilty rather than judges being racial biased (Lammy, 2017). Third, racial disproportionality in the CJS might be the consequence of the intersectionality of many disadvantages experienced by certain BAME groups, including them also having other social and political identities that correlate with an increased likelihood of overrepresentation in the CJS e.g., being of low socioeconomic status, poor

and working class (Crenshaw, 1989). Consequently, it is perhaps the accumulation of these social and political identities that contribute to the overrepresentation of certain BAME people in the CJS, rather than the single factor of being of a specific BAME group.

To conclude, in some respects the finding from the systematic review of inconsistent evidence of racial bias by mock jurors, alongside the non-significant findings in the empirical studies, might be seen as evidence that generally CJWs and (mock) jurors are not racially biased and are functioning well. However, the current thesis only examined two aspects of the CJS, which is whether mock jurors and CJWs are racially biased in their decisions on guilt, blame and punishment. Perhaps racial bias is occurring elsewhere in the CJS, such as at the point of arrest or decisions to prosecute. Amongst the several recommendations already mentioned in this chapter, future research needs to also investigate how decisions made at other stages of the CJS might account for racial disproportionality.

References

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk Research: Review and Recommendations. *Journal of management, 47*(4), 823-837. doi:10.1177/0149206320969787

Andrade, C. (2019). The P value and statistical significance: misunderstandings, explanations, challenges, and alternatives. *Indian Journal of Psychological Medicine, 41*(3), 210-215.

Arifin, W. N. (2012). Random sampling and allocation using SPSS. *Education in Medicine Journal, 4*(1).

Armenta, A. D. (2017). *Mothers Who Kill: An Aversive Bias Examination of Juror Prejudice Against Mothers Accused of Filicide.* (10604928 M.A.). California State University, Fullerton, Ann Arbor. Retrieved from https://www.proquest.com/openview/7fe5e0b1275b34447b90eee3e74ca062/1?pq-origsite=gscholar&cbl=18750

https://nusearch.nottingham.ac.uk/openurl/44NOTUK/44NOTUK?genre=dissertations+%26+theses&atitle=&author=Armenta%2C+Angel+David&volume=&issue=&spage=&date=2017-01-01&rft.btitle=&rft.jtitle=&issn=&isbn=978-0-355-16435-0&sid=ProQuest+Dissertations+%26+Theses+A%26I_ ProQuest Dissertations & Theses A&I database.

Asgeirsdottir, R. L., Vésteinsdóttir, V., & Thorsdottir, F. (2016). Short form development of the Balanced Inventory of Desirable Responding: Applying confirmatory factor analysis, item response theory, and cognitive interviews to scale reduction. *Personality and Individual Differences, 96*, 212-221.

Australian Bureau of Statistics (ABS) (2017). http://www.abs.ov.au/ausstats/abs@nsf/mf/4512.0

Axt, J. R., Ebersole, C. R., & Nosek, B. A. (2016). An unintentional, robust, and replicable pro-Black bias in social judgment. *Social Cognition, 34*(1), 1-39. doi:10.1521/soco.2016.34.1.1

Ayres, I., & Waldfogel, J. (1993). A market test for race discrimination in bail setting. *Stanford Law Review, 46*, 987.

Baldus, D. C., Woodworth, G., Zuckerman, D., Weiner, N. A., & Broffitt, B. (1998). Racial discrimination and the death penalty in the post-Furman era: an empirical and legal overview, with recent findings from Philadelphia. *Cornell Law Review, 83*(6), 1638-1770.

Barchard, K. (2002). *The discriminant validity of positive expressivity and negative expressivity.* Paper presented at the Poster presented at the International Society for Research on Emotions

Annual Convention, July.

Barger, S. D. (2002). The Marlowe-Crowne affair: Short forms, psychometric structure, and social desirability. *Journal of Personality Assessment, 79*(2), 286-305.

Barry, M. (2010). Youth transitions: From offending to desistance. *Journal of Youth Studies, 13*(1), 121-136.

Barter, C., & Renold, E. (2000). 'I wanna tell you a story': exploring the application of vignettes in qualitative research with children and young people. *International Journal of Social Research Methodology, 3*(4), 307-323.

Bartol, C. R., & Bartol, A. M. (1994). Psychology and law. *Pacific Grove, CA: Brooks/Cole*.

Beckett, N. E., & Park, B. (1995). Use of category versus individuating information: Making base rates salient. *Personality and Social Psychology Bulletin, 21*(1), 21-31.

Bell, C. (2019). *If Looks Could Kill: Does Female Attractiveness Influence Death Penalty Decision Making.* (13813435 M.A.). University of West Georgia, Ann Arbor. Retrieved from http://nottingham.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/if-looks-could-kill-does-female-attractiveness/docview/2234281123/se-2?accountid=8018

Benedict, C., Hahn, A. L., Diefenbach, M. A., & Ford, J. S. (2019). Recruitment via social media: advantages and potential biases. *Digital Health, 5*, 2055207619867223.

Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement, 62*(4), 570-589.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political analysis, 20*(3), 351-368. doi:10.1093/pan/mpr057

Binder, J., Zagefka, H., Brown, R., Funke, F., Kessler, T., Mummendey, A., Leyens, J. P. (2009). Does contact reduce prejudice or does prejudice reduce contact? A longitudinal test of the contact hypothesis among majority and minority groups in three European countries. *Journal of Personality and Social Psychology, 96*(4), 843.

Bjerregaard, B. E., Smith, M. D., Cochran, J. K., & Fogel, S. J. (2017). A further examination of the liberation hypothesis in capital murder trials. *Crime & Delinquency, 63*(8), 1017-1038.

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in

criminal sentencing. *Psychological Science*, *15*(10), 674-679.

Bobbio, A., & Manganelli, A. M. (2011). Measuring social desirability responding. A short version of Paulhus' BIDR 6. *Testing, Psychometrics Methodology in Applied Psychology, 18*(2), 117-135.

Bodenhausen, G. V., & Lichtenstein, M. (1987). Social stereotypes and information-processing strategies: the impact of task complexity. *Journal of Personality and Social Psychology, 52*(5), 871.

Bodenhausen, G. V., & Wyer, R. S. (1985). Effects of stereotypes in decision making and information-processing strategies. *Journal of Personality and Social Psychology, 48*(2), 267.

Borg, M. J. (1997). The Southern Subculture of Punitiveness? Regional Variation in Support for Capital Punishment. *The Journal of Research in Crime and Delinquency, 34*(1), 25-45. doi:10.1177/0022427897034001003

Bornstein, B. H., Golding, J. M., Neuschatz, J., Kimbrough, C., Reed, K., Magyarics, C., & Luecht, K. (2017). Mock Juror Sampling Issues in Jury Simulation Research: A Meta-Analysis. *Law and human behavior, 41*(1), 13-28. doi:10.1037/lhb0000223

Bottoms, B. L., Davis, S. L., & Epstein, M. A. (2004). Effects of Victim and Defendant Race on Jurors' Decisions in Child Sexual Abuse Cases. *Journal of Applied Social Psychology, 34*(1), 1-33. doi:10.1111/j.1559-1816.2004.tb02535.

Bowers, W. J., Steiner, B. D., & Sandys, M. (2001). Death sentencing in Black and White: An empirical analysis of the role of jurors' race and jury racial composition. *University of Pennsylvania Journal of Constitutional Law., 3*, 171.

Bradbury-Jones, C., Taylor, J., & Herber, O. (2014). How theory is used and articulated in qualitative research: Development of a new typology. *Social Science & Medicine, 120*, 135-141.

Bridges, G. S., Crutchfield, R. D., & Simpson, E. E. (1987). Crime, social structure and criminal punishment: White and non-White rates of imprisonment. *Social Problems, 34*(4), 345-361.

Bridges, G. S., & Steen, S. (1998). Racial disparities in official assessments of juvenile offenders: Attributional stereotypes as mediating mechanisms. *American Sociological Review*, 554-570.

Bullock, H. A. (1961). Significance of the racial factor in the length of prison sentences. *J. Crim. L. Criminology & Police Sci., 52*, 411.

Burke, K. C. (2019). *Does Collective Guilt Minimize Racial Bias in Jurors' Decisions?* (27605401 M.A.). University of Illinois at Chicago, Ann Arbor. Retrieved from http://nottingham.idm.oclc.org/login?url=https://www.proquest.com/dissertations-

theses/does-collective-guilt-minimize-racial-bias-jurors/docview/2331474726/se-2?accountid=8018

https://nusearch.nottingham.ac.uk/openurl/44NOTUK/44NOTUK?genre=dissertations+%26+theses&atitle=&author=Burke%2C+Kelly+C.&volume=&issue=&spage=&date=2019-01-01&rft.btitle=&rft.jtitle=&issn=&isbn=9781687985743&sid=ProQuest+Dissertations+%26+Theses+A%26I_ ProQuest Dissertations & Theses A&I database.

Calabrese, S. K., Earnshaw, V. A., Magnus, M., Hansen, N. B., Krakower, D. S., Underhill, K., . . . Dovidio, J. F. (2018). Sexual stereotypes ascribed to Black men who have sex with men: An intersectional analysis. *Archives of Sexual Behavior, 47*(1), 143-156.

Call, A. A. (2019). The effects of child race, child age, and defendant race on mock jurors' decisions for a child sexual abuse case. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 80*(11-B(E)), No-Specified. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc16&NEWS=N&AN=2019-46356-201

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105. doi:10.1037/h0046016

Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods*, *15*(1), 17-32.

Carnes, M., Devine, P. G., Baier Manwell, L., Byars-Winston, A., Fine, E., Ford, C. E., ... Sheridan, J. (2015). The effect of an intervention to break the gender bias habit for faculty at one institution: A cluster randomized, controlled trial. *Academic medicine: journal of the Association of American Medical Colleges*, *90*(2), 221.

Cervellione, K. L., Lee, Y.-S., & Bonanno, G. A. (2009). Rasch modeling of the self-deception scale of the balanced inventory of desirable responding. *Educational and Psychological Measurement, 69*(3), 438-458.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5), 752.

Charman, S. D., Kavetski, M., & Mueller, D. H. (2017). Cognitive bias in the legal system: Police officers evaluate ambiguous evidence in a belief-consistent manner. *Journal of Applied Research in Memory and Cognition, 6*(2), 193-202.

Chmielewski, M., & Kucker, S. C. (2020). An MTurk Crisis? Shifts in Data Quality and the Impact on

Study Results. *Social Psychological and Personality Scienc, 11*(4), 464-473. doi:10.1177/1948550619875149

Chyung, S. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement, 56*(10), 15-23.

Corp, I. (2013). IBM SPSS statistics for windows, version 22.0. *Armonk, NY: IBM Corp*.

Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology, 92*(6), 1006.

Cox, W. T., & Devine, P. G. (2015). Stereotypes possess heterogeneous directionality: A theoretical and empirical exploration of stereotype structure and content. *PloS one, 10*(3), e0122292.

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 139.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Crowley, M. J., O'Callaghan, M. G., & Ball, P. J. (1994). The juridical impact of psychological expert testimony in a simulated child sexual abuse trial. *Law and human behavior, 18*(1), 89-105.

Crowne, D., & Marlowe, D. (1964). *The Approval Motive*. Wiley, New York.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349.

Crutzen, R., & Göritz, A. S. (2010). Social desirability and self-reported health risk behaviors in web-based research: three longitudinal studies. *BMC public health, 10*(1), 1-10.

Dahiru, T. (2008). P-value, a true test of statistical significance? A cautionary note. *Annals of Ibadan postgraduate medicine, 6*(1), 21-26.

Damen, N. L., Versteeg, H., Serruys, P. W., van Geuns, R.-J. M., Van Domburg, R. T., Pedersen, S. S., & Boersma, E. (2015). Cardiac patients who completed a longitudinal psychosocial study had a different clinical and psychosocial baseline profile than patients who dropped out prematurely. *European Journal of Preventive Cardiology, 22*(2), 196-199.

Dane, F., & Wrightsman, L. (1983). Effects of Defendants' and Victims' Characteristics on Jurors' Verdicts (From Psychology of the Courtroom, P 83-115, 1982, Norbert L Kerr and Robert M Bray, ed.-See NCJ-89761).

David, M. M. (2016). *Influence of metaphoric language on juror decision making.* (10172401 M.A.). Loyola University Chicago, Ann Arbor. Retrieved from http://nottingham.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/influence-metaphoric-language-on-juror-decision/docview/1849497627/se-2?accountid=8018

https://nusearch.nottingham.ac.uk/openurl/44NOTUK/44NOTUK?genre=dissertations+%26+theses&atitle=&author=David%2C+Melissa+M.&volume=&issue=&spage=&date=2016-01-01&rft.btitle=&rft.jtitle=&issn=&isbn=978-1-369-27167-6&sid=Psychology+Database_ Psychology Database database.

Deaux, K. (1976). *The behavior of women and men*: Monterey, CA: Brooks/Cole.

Deaux, K. (1976). Sex: A perspective on the attribution process. *New Directions in Attribution Research, 1*, 335-352.

DeCesare, B. (2017). *Effects of Defendant Group Status and Inclusion of a Lesser Charge on Participant-Juror Verdict Preferences.* (10286582 M.A.). City University of New York John Jay College of Criminal Justice, Ann Arbor. Retrieved from http://nottingham.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/effects-defendant-group-status-inclusion-lesser/docview/1914912937/se-2?accountid=8018

https://nusearch.nottingham.ac.uk/openurl/44NOTUK/44NOTUK?genre=dissertations+%26+theses&atitle=&author=DeCesare%2C+Brittany&volume=&issue=&spage=&date=2017-01-01&rft.btitle=&rft.jtitle=&issn=&isbn=978-1-369-82179-6&sid=ProQuest+Dissertations+%26+Theses+A%26I_ ProQuest Dissertations & Theses A&I database.

DeVellis R. (2003). *Scale development: theory and applications: theory and application*. Thousand Oaks, CA: Sage.

Devine, D. J., & Caughlin, D. E. (2014). Do they matter? A meta-analytic investigation of individual characteristics and guilt judgments. *Psychology, Public Policy, and Law, 20*(2), 109.

Devine, P. G., P. S. Forscher, A. J. Austin and W. T. L. Cox. 2012. 'Long-Term Reduction in Implicit Race Bias: A Prejudice Habit-Breaking Intervention.' *Journal of Experimental Social Psychology 48*(6):1267–1278.

Dilchert. S., & One, D. (2011). Application of prevention strategies. In Ziegler, M., MacCann, C., & Roberts, R. (Eds.). *New Perspectives on Faking in Personality Assessment*. Oxford University Press.

Dillman, D. (2000). Constructing the questionnaire. Mail and internet surveys. In: New York: John Wiley & Sons.

Dillman, D. A. (2011). *Mail and Internet surveys: The tailored design method--2007 Update with new Internet, visual, and mixed-mode guide*: John Wiley & Sons.

Dodson, P. (2015). *Hidden risk factors: Do race and biological sex influence perceived culpability and decisions made about juveniles facing waiver to adult court?* (3704431 Psy.D.). Alliant International University, Ann Arbor. Retrieved from http://nottingham.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/hidden-risk-factors-do-race-biological-sex/docview/1690497474/se-2?accountid=8018

https://nusearch.nottingham.ac.uk/openurl/44NOTUK/44NOTUK?genre=dissertations+%26+theses&atitle=&author=Dodson%2C+Porsha&volume=&issue=&spage=&date=2015-01-01&rft.btitle=&rft.jtitle=&issn=&isbn=978-1-321-76823-7&sid=Criminal+Justice+Database_ Criminal Justice Database; Social Science Premium Collection database.

Douglas, E. F., McDaniel, M. A., & Snell, A. F. (1996, August). The validity of non-cognitive measures decays when applicants fake. In *Academy of Management Proceedings* (Vol. 1996, No. 1, pp. 127-131). Briarcliff Manor, NY 10510: Academy of Management.

Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological science, 17*(5), 383-386.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden Press.

Edwards, A. L. (1970). *The measurement of personality traits by scales and inventories*. New York: Holt, Rinehart & Winston.

Edwards, A. L., & Diers, C. J. (1962). Social desirability and the factorial interpretation of the MMPI. *Educational and Psychological Measurement, 22*(3), 501-509.

Ellis, D. M. (2020). *Juveniles Tried as Adults: The Impact of Youth Demographic Factors on Juror Perceptions.* (28149833 M.A.). City University of New York John Jay College of Criminal Justice, Ann Arbor. Retrieved from http://nottingham.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/juveniles-tried-as-adults-impact-youth/docview/2508898780/se-2?accountid=8018

Else-Quest, N. M., & Hyde, J. S. (2016). Intersectionality in quantitative psychological research: I. Theoretical and epistemological issues. *Psychology of women quarterly, 40*(2), 155-170.

Espinoza, R. K. E., Willis-Esqueda, C., Toscano, S., & Coons, J. V. (2015). The impact of ethnicity, immigration status and SES on Juror decision making. *Journal of Ethnicity in Criminal Justice, 13* (3), 197–216. https://doi.org/10.1080/15377938.2014.984044

Esqueda, C. W., Espinoza, R. K., & Culhane, S. E. (2008). The effects of ethnicity, SES, and crime status on juror decision making: A cross-cultural examination of European American and Mexican American mock jurors. *Hispanic Journal of Behavioral Sciences*, *30*(2), 181-199.

Eysenck, H. J., & Eysenck, S. (1965). The Eysenck personality inventory. *British Journal of Educational Studies, 14*(1).

Federal Bureau of Prisons. (2019). BOP: Statistics. Retrieved September 28, 2019, from http://www.bop.gov/about/statistics/.

Feild, H. S. (1979). Rape Trials and Jurors' Decisions: A Psycholegal Analysis of the Effects of Victim, Defendant, and Case Characteristics. *Law and Human Behavior, 3*(4), 261-284. doi:10.1007/BF01039806

Ferreira, J. (2021). Problems of Contemporary Society: How Terrorism Shaped Today's Britain. *E-Revista de Estudos Interculturais, 9*(1).

Finch, J. (1987). The vignette technique in survey research. *Sociology, 21*(1), 105-114.

Findlay, M., & Duff, P. (1987). *The jury under attack*. London & Sydney: Butterworths.

Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe-Crowne social desirability scale. *Educational and psychological measurement, 53*(2), 417-424.

Fiske, S. T., & Taylor, S. E. (1991). Social cognition. McGraw-Hill. *New York*.

FitzGerald, C., Martin, A., Berner, D., & Hurst, S. (2019). Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review. *BMC Psychology*, *7*(1), 1-12.

Foley, L. A., & Chamblin, M. H. (1982). The Effect of Race and Personality on Mock Jurors' Decisions.

The Journal of Psychology, 112(1), 47-51. doi:10.1080/00223980.1982.9923533

Forscher, P. S., Mitamura, C., Dix, E. L., Cox, W. T., & Devine, P. G. (2017). Breaking the prejudice habit: Mechanisms, timecourse, and longevity. *Journal of experimental social psychology, 72,* 133-146.

ForsterLee, R., ForsterLee, L., Horowitz, I. A., & King, E. (2006). The effects of defendant race, victim race, and juror gender on evidence processing in a murder trial. *Behavioral Sciences & the Law, 24*(2), 179-198.

Franklin, T. W., & Fearn, N. E. (2015). Sentencing Asian offenders in state courts: The influence of a prevalent stereotype. *Crime & Delinquency*, *61*(1), 96-120.

Frenkel-Brunswik, E. (1939). Mechanisms of self-deception. *The Journal of Social Psychology, 10*(3), 407-420.

Gabora, N. J., Spanos, N. P., & Joab, A. (1993). The effects of complainant age and expert psychological testimony in a simulated child sexual abuse trial. *Law and Human Behavior, 17*(1), 103-119.

Gamblin, B. W., & Kehn, A. (2021). Race salience and attorney statements: the unique role of defense opening statements and closing arguments: Research and Reviews. *Current Psychology, 40*(6), 2621-2633. doi:http://dx.doi.org/10.1007/s12144-020-01147-8

Garton L, Haythornthwaite C, Wellman B (1999) Studying On-Line Social Networks. Doing Internet Research: Critical Issues and Methods for Examining the Net 3, pp 75–105

George, W. H., & Martínez, L. J. (2002). Victim Blaming in Rape: Effects of Victim and Perpetrator Race, Type of Rape, and Participant Racism. *Psychology of Women Quarterly, 26*(2), 110-119. doi:10.1111/1471-6402.00049

Glaser, J., Martin, K. D., & Kahn, K. B. (2015). Possibility of death sentence has divergent effect on verdicts for Black and White defendants. *Law and Human Behavior, 39*(6), 539-546. doi:http://dx.doi.org/10.1037/lhb0000146

Goodman-Delahunty, J., & Sporer, S. L. (2010). Unconscious influences in sentencing decisions: A research review of psychological sources of disparity. *Australian Journal of Forensic Sciences, 42*(1), 19-36.

Gordon, R. A. (1990). Attributions for Blue-Collar and White-Collar Crime: The Effects of Subject and

Defendant Race on Simulated Juror Decisions 1. *Journal of Applied Social Psychology, 20*(12), 971-983.

Gordon, R. A. (1993). The Effect of Strong Versus Weak Evidence on the Assessment of Race Stereotypic and Race Nonstereotypic Crimes 1. *Journal of Applied Social Psychology, 23*(9), 734-749.

Gordon, R. A., Bindrim, T. A., McNicholas, M. L., & Walden, T. L. (1988). Perceptions of Blue-Collar and White-Collar Crime: The Effect of Defendant Race on Simulated Juror Decisions. *The Journal of Social Psychology, 128*(2), 191-197. doi:10.1080/00224545.1988.9711362

Government of Canada, Office of Correctional Investigator (OCI) (2016). http://www.oci-bec-gc.ca/cnt/rpt/annrpt/annrpt20152016-eng.aspx

Graham, S., & Lowery, B. S. (2004). Priming unconscious racial stereotypes about adolescent offenders. *Law and human behavior, 28*(5), 483-504.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement, 37*(4), 827–838.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337-350.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4.

Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California law review, 94*(4), 945-967.

Haegerich, T. M., Salerno, J. M., & Bottoms, B. L. (2013). Are the effects of juvenile offender stereotypes maximized or minimized by jury deliberation? *Psychology, Public Policy, and Law, 19*(1), 81-97. doi:10.1037/a0027808

Hans, V. P., & Vidmar, N. (2013). *Judging the jury*: Springer.

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: a meta-analysis of 118 prediction studies. *Psychological assessment, 21*(1), 1.

Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The balanced inventory of desirable responding short form (BIDR-16). *SAGE Open, 5*(4), 2158244015621113.

Hart, S. D. (2016). Culture and violence risk assessment: The case of Ewert v. Canada. *Journal of Threat Assessment and Management, 3*(2), 76.

Hathaway, S. R., & McKinley, J. C. (1951). Minnesota multiphasic personality inventory; manual, revised.

Hilbert, S., Kuechenhoff, H., Sarubin, N., TOYO NAKAGAWA, T., & Buehner, M. (2016). The influence of the response format in a personality questionnaire: An analysis of a dichotomous, a Likert-type, and a visual analogue scale. *TPM: Testing, Psychometrics, Methodology in Applied Psychology, 23*(1).

Holden, R. R. (2007). Socially desirable responding does moderate personality scale validity both in experimental and in nonexperimental contexts. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 39*(3), 184.

Holden, R. R. (2008). Underestimating the effects of faking on the validity of self-report personality scales. *Personality and Individual Differences, 44*(1), 311-321.

Holden, R. R., Book, A. S., Edwards, M. J., Wasylkiw, L., & Starzyk, K. B. (2003). Experimental faking in self-reported psychopathology: unidimensional or multidimensional? *Personality and Individual Differences, 35*(5), 1107-1117.

Holden, R. R., Starzyk, K. B., McLeod, L. D., & Edwards, M. J. (2000). Comparisons among the holden psychological screening inventory (HPSI), the brief symptom inventory (BSI), and the balanced inventory of desirable responding (BIDR). *Assessment, 7*(2), 163-175.

Home Office. (2020, July 30). Statistics at Home Office. *https://www.gov.uk/government/organisations/home-office/about/statistics*. Retrieved from https://www.gov.uk/government/statistics/police-workforce-england-and-wales-31-march-2020

Horst, P. (1968). *Personality; measurement of dimensions*. San Francisco: Jossey-Bass.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

Hubley, A. M. (2014). Discriminant validity. *Encyclopedia of quality of life and well-being research*, 1664-1667.

Huff, C., & Tingley, D. (2015). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & politics, 2*(3), 205316801560464. doi:10.1177/2053168015604648

Hughes, R., & Huby, M. (2004). The construction and interpretation of vignettes in social research.

Social work and social sciences review, 11(1), 36-51.

Hymes, R. W., Leinart, M., Rowe, S., & Rogers, W. (1993). Acquaintance rape: The effect of race of defendant and race of victim on White juror decisions. *The Journal of Social Psychology, 133*(5), 627-634.

Ilieva, J., Baron, S., & Healey, N. M. (2002). Online surveys in marketing research. *International Journal of Market Research, 44*(3), 1-14.

Jackson, D. N., & Lay, C. H. (1968). Homogeneous dimensions of personality scale content. *Multivariate Behavioral Research, 3*(3), 321-337.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality Theory and Research, ed Pervin L-A (Guilford, New York), pp 102–138*.

Johnson, B. D., & Betsinger, S. (2009). Punishing the "model minority": Asian-American criminal sentencing outcomes in federal district courts. *Criminology*, *47*(4), 1045-1090.

Johnson, B. D., Van Wingerden, S., & Nieuwbeerta, P. (2010). Sentencing homicide offenders in the Netherlands: Offender, victim, and situational influences in criminal punishment. *Criminology, 48*(4), 981-1018.

Johnson, J. D., Whitestone, E., Jackson, L. A., & Gatto, L. (1995). Justice is still not colorblind: Differential racial effects of exposure to inadmissible evidence. *Personality and Social Psychology Bulletin, 21*(9), 893-898.

Johnson, T. P., Fendrich, M., & Hubbell, A. (2002). *A validation of the Crowne-Marlowe social desirability scale.* Paper presented at the 57th Annual meeting of the American Association for Public Opinion Research.

Jones, C. S., & Kaplan, M. F. (2003). The effects of racially stereotypical crimes on juror decision-making and information-processing strategies. *Basic and Applied Social Psychology, 25*(1), 1-13.

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76*(5), 349.

Kalven, H., Zeisel, H., Callahan, T., & Ennis, P. (1966). *The American Jury*: Little, Brown Boston.

Kam, C. (2013). Probing item social desirability by correlating personality items with Balanced Inventory of Desirable Responding (BIDR): A validity examination. *Personality and Individual Differences, 54*(4), 513-518.

Kang, J., & Banaji, M. R. (2006). Fair measures: A behavioral realist revision of affirmative action. *California Law Review., 94*, 1063.

Kang, S. K., & Bodenhausen, G. V. (2015). Multiple identities in social perception and interaction: Challenges and opportunities. *Annual review of psychology, 66*, 547-574.

Kapardis, A. (1997). *Psychology and law: A critical appraisal*. Cambridge University Press.

Kaplan, M. F., & Miller, L. E. (1978). Effects of jurors' identification with the victim depend on likelihood of victimization. *Law and Human Behavior, 2*(4), 353-361.

Kaplan, M. F., & Miller, L. E. (1978). Reducing the effects of juror bias. *Journal of Personality and Social Psychology, 36*(12), 1443.

Keil, T. J., & Vito, G. F. (1995). Race and the death penalty in Kentucky murder trials: 1976–1991. *American Journal of Criminal Justice, 20*(1), 17-36.

Kennedy, R. (2012). *Race, crime, and the law*: Vintage.

Kepes, S., McDaniel, M. A., Brannick, M. T., & Banks, G. C. (2013). Meta-analytic reviews in the organizational sciences: Two meta-analytic schools on the way to MARS (the Meta-Analytic Reporting Standards). *Journal of Business and Psychology*, *28*(2), 123-143.

Kerr, N. L., & Bray, R. M. (Eds.). (1982). *The psychology of the courtroom*. New York: Academic Press.

Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996). Bias in Judgment: Comparing Individuals and Groups. *Psychological Review, 103*(4), 687-719. doi:10.1037/0033-295X.103.4.687

King, R. D., & Light, M. T. (2019). Have racial and ethnic disparities in sentencing declined? *Crime and Justice, 48*(1), 365-437.

Klein, K., & Creech, B. (1982). Race, Rape, and Bias: Distortion of Prior Odds and Meaning Changes. *Basic and Applied Social Psychology, 3*(1), 21-33. doi:10.1207/s15324834basp0301_2

Klein, S. P., & Rolph, J. E. (1991). Relationship of offender and victim race to death penalty sentences in California. *Jurimetrics J., 32*, 33.

Kluwe, K. (2015). *The influence of racial and crime stereotypes on jurors' representations of trial evidence and verdict decisions.* (3713583 Ph.D.). Loyola University Chicago, Ann Arbor. Retrieved from http://nottingham.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/influence-racial-crime-stereotypes-on-jurors/docview/1708673527/se-2?accountid=8018

https://nusearch.nottingham.ac.uk/openurl/44NOTUK/44NOTUK?genre=dissertations+%26+theses&

atitle=&author=Kluwe%2C+Katharina&volume=&issue=&spage=&date=2015-01-01&rft.btitle=&rft.jtitle=&issn=&isbn=978-1-321-90175-7&sid=Psychology+Database_ Psychology Database database.

Kros, M., & Hewstone, M. (2020). Negative and positive interethnic contact and the Association of Ethnic Neighbourhood Composition with trust, cohesion, and prejudice. *European Sociological Review, 36*(6), 937-956.

Krueger, J., & Clement, R. W. (1994). The truly false consensus effect: an ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology, 67*(4), 596.

Kutateladze, B. L., Andiloro, N. R., Johnson, B. D., & Spohn, C. C. (2014). Cumulative disadvantage: Examining racial and ethnic disparity in prosecution and sentencing. *Criminology*, *52*(3), 514-551.

Kurinec, C. A., & Weaver, C. A., III. (2019). Dialect on trial: use of African American Vernacular English influences juror appraisals. *Psychology, Crime & Law : PC & L, 25*(8), 803-828. doi:http://dx.doi.org/10.1080/1068316X.2019.1597086

Lammy, D. (2017). The Lammy Review: An independent review into the treatment of, and outcomes for, Black, Asian and Minority Ethnic individuals in the Criminal Justice System. *London: Lammy Review*.

Landy, D., & Aronson, E. (1969). The influence of the character of the criminal and his victim on the decisions of simulated jurors. *Journal of Experimental Social Psychology, 5*(2), 141-152.

Lee, C. K. (1997). Race and the victim: An examination of capital sentencing and guilt attribution studies. *Chicago-Kent Law Review, 73*, 533.

Lee, E. (2015). *The making of Asian America: A history*. New York, NY: Simon & Schuster.

Leifker, D., & Sample, L. L. (2011). Probation recommendations and sentences received: The association between the two and the factors that affect recommendations. *Criminal Justice Policy Review, 22*(4), 494-517.

Leippe, M. R., Bergold, A. N., & Eisenstadt, D. (2017). Prejudice and terror management at trial: Effects of defendant race/ethnicity and mortality salience on mock-jurors' verdict judgments. *The Journal of Social Psychology, 157*(3), 279-294. doi:http://dx.doi.org/10.1080/00224545.2016.1184128

Leite, W. L., & Beretvas, S. N. (2005). Validation of scores on the Marlowe-Crowne social desirability scale and the balanced inventory of desirable responding. *Educational and psychological*

measurement, 65(1), 140-154.

Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment, 14*(2), 131-141.

Linden, W., Paulhus, D. L., & Dobson, K. S. (1986). Effects of response styles on the report of psychological and somatic distress. *Journal of Consulting and Clinical Psychology, 54*(3), 309.

Lönnqvist, J. E., Paunonen, S., Tuulio-Henriksson, A., Lönnqvist, J., & Verkasalo, M. (2007). Substance and style in socially desirable responding. *Journal of Personality, 75*(2), 291-322.

Lowder, E. M., Desmarais, S. L., Rade, C. B., Johnson, K. L., & Van Dorn, R. A. (2019). Reliability and validity of START and LSI-R assessments in mental health jail diversion clients. *Assessment, 26*(7), 1347-1361.

Lynch, M., & Haney, C. (2000). Discrimination and instructional comprehension: Guided discretion, racial bias, and the death penalty. *Law and Human Behavior*, *24*(3), 337-358.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19.

Macpherson, W. (1999). *The stephen lawrence inquiry* (Vol. 1): Stationery Office Limited London.

Maeder, E., & Yamamoto, S. (2019). Social Identity in the Canadian Courtroom: Effects of Juror and Defendant Race. *Canadian Journal of Criminology and Criminal Justice, 61*(4), 24-24–44. doi:http://dx.doi.org/10.3138/cjccj.2018-0057

Maeder, E. M., & Logan, E. (2018). What Makes Race Salient? Juror Decision-Making in Same-Race Versus Cross-Race Identification Scenarios and the Influence of Expert Testimony. *Criminal Justice and Behavior, 45*(8), 1234-1251. doi:http://dx.doi.org/10.1177/0093854818776998

Maeder, E. M., & McManus, L. A. (2020). Mosaic or Melting Pot? Race and Juror Decision Making in Canada and the United States. *Journal of Interpersonal Violence*. doi:10.1177/0886260520917512

Maeder, E. M., & Yamamoto, S. (2019). Investigating Race Salience, Defendant Race, and Victim Race Effects on Mock Juror Decision-Making in Canada. *Justice Quarterly : JQ, 36*(5), 929-953. doi:http://dx.doi.org/10.1080/07418825.2018.1460390

Maeder, E. M., Yamamoto, S., & McLaughlin, K. J. (2020). The influence of defendant race and mental disorder type on mock juror decision-making in insanity trials. *International Journal of Law and Psychiatry, 68*. doi:10.1016/j.ijlp.2019.101536

Maeder, E. M., Yamamoto, S., & McManus, L. A. (2015). Race salience in Canada: Testing multiple manipulations and target races. *Psychology, Public Policy, and Law, 21*(4), 442-451. doi:http://dx.doi.org/10.1037/law0000057

Maeder, E. M., Yamamoto, S., McManus, L. A., & Capaldi, C. A. (2016). Race−crime congruency in the Canadian context. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 48*(2), 162-170. doi:http://dx.doi.org/10.1037/cbs0000045

Maeder, E. M., Yamamoto, S., & Saliba, P. (2015). The influence of defendant race and victim physical attractiveness on juror decision-making in a sexual assault trial. *Psychology, Crime & Law : PC & L, 21*(1), 62. Retrieved from http://nottingham.idm.oclc.org/login?url=https://www.proquest.com/scholarly-journals/influence-defendant-race-victim-physical/docview/1628012608/se-2?accountid=8018

Marcus-Newhall, A., Blake, L. P., & Baumann, J. (2002). Perceptions of hate crime perpetrators and victims as influenced by race, political orientation, and peer group. *American Behavioral Scientist, 46*(1), 108-135.

Marques, J. M., Yzerbyt, V. Y., & Leyens, J. P. (1988). The "Black sheep effect": Extremity of judgments towards ingroup members as a function of group identification. *European journal of social psychology, 18*(1), 1-16.

Mazzella, R., & Feingold, A. (1994). The Effects of Physical Attractiveness, Race, Socioeconomic Status, and Gender of Defendants and Victims on Judgments of Mock Jurors: A Meta-Analysis 1. *Journal of Applied Social Psychology, 24*(15), 1315-1338.

McIntire, S. A., & Miller, L. A. (2000). *Foundations of psychological testing. New York, NY*: McGraw-Hill.

McKinley, J., Hathaway, S., & Meehl, P. (1948). The Minnesota Multiphasic Personality Inventory: VI. The K Scale. *Journal of Consulting Psychology, 12*(1), 20.

McManus, L., Maeder, E., & Yamamoto, S. (2018). The Role of Defendant Race and Racially Charged Media in Canadian Mock Juror Decision Making. *Canadian Journal of Criminology and Criminal Justice, 60*(2), 266. Retrieved from http://nottingham.idm.oclc.org/login?url=https://www.proquest.com/scholarly-journals/role-defendant-race-racially-charged-media/docview/2075655077/se-2?accountid=8018

Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology, 30*(5), 525.

Miller, D. T., & Turnbull, W. (1986). Expectancies and interpersonal processes. *Annual Review of Psychology, 37*(1), 233-256.

Millham, J., & Jacobson, L. I. (1978). The need for approval. *Dimensions of personality*, 365-390.

Mitchell, T. L., Haw, R. M., Pfeifer, J. E., & Meissner, C. A. (2005). Racial bias in mock juror decision-making: A meta-analytic review of defendant treatment. *Law and Human Behavior*, *29*(6), 621-637. doi:10.1007/s10979-005-8122-9

Montag, I., & Comrey, A. L. (1982). Personality construct similarity in Israel and the United States. *Applied Psychological Measurement, 6*(1), 61-67.

Murrie, D. C., Boccaccini, M. T., Turner, D. B., Meeks, M., Woods, C., & Tussey, C. (2009). Rater (dis) agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and Law, 15*(1), 19.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European journal of social psychology, 15*(3), 263-280.

Nesbitt, B. N. (2019). *Race and Ethnicity Influences on Mock-Juror Evaluations of Eyewitness Testimony.* (27662902 M.A.). Baylor University, Ann Arbor. Retrieved from http://nottingham.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/race-ethnicity-influences-on-mock-juror/docview/2352130581/se-2?accountid=8018

Nunan, D., Aronson, J., & Bankhead, C. (2018). Catalogue of bias: attrition bias. *BMJ evidence-based medicine, 23*(1), 21-22.

Nygren, L., & Oltedal, S. (2015). Constructing a vignette for qualitative comparative family research. *Journal of Comparative Social Work, 10*(1), 73-76.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*(6), 660.

Paddam, A., Barnes, D., & Langdon, D. (2010). Constructing vignettes to investigate anger in multiple sclerosis. *Nurse researcher, 17*(2).

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current directions in psychological science: a journal of the American Psychological Society, 23*(3), 184-188. doi:10.1177/0963721414531598

Paternoster, R., & Brame, R. (2008). Reassessing race disparities in Maryland capital cases.

Criminology, 46(4), 971-1008.

Paulhus, D. (1988). Balanced inventory of desirable responding (BIDR). *Acceptance and Commitment Therapy. Measures Package, 41*, 79586-79587.

Paulhus, D. L. (1981). Control of social desirability in personality inventories: Principal-factor deletion. *Journal of Research in Personality, 15*(3), 383-388.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598.

Paulhus, D. L. (1991). Measurement and control of response bias.

Paulhus, D. L. (1994). Balanced inventory of desirable responding: Reference manual for BIDR version 6. *Unpublished manuscript, University of British Columbia, Vancouver, Canada*.

Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing? *Journal of Personality and Social Psychology, 74*(5), 1197.

Paulhus, D. L. (1998). *Paulhus deception scales (PDS): the balanced inventory of desirable responding-7: user's manual.* North Tanawanda, NY: Multi-Health Systems.

Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin, 21*(2), 100-108.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*(6), 1025-1060.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*(2), 307.

Paulhus, D. L., & Vazire, S. (2007). The self-report method. *Handbook of Research Methods in Personality Psychology, 224–239.*

Paunonen, S. V., & LeBel, E. P. (2012). Socially desirable responding and its elusive effects on the validity of personality assessments. *Journal of Personality and Social Psychology, 103*(1), 158.

Payne, B. K., & Chappell, A. (2008). Using Student Samples in Criminological Research. *Journal of criminal justice education, 19*(2), 175-192. doi:10.1080/10511250802137226

Perrault, R. T., Vincent, G. M., & Guy, L. S. (2017). Are risk assessments racially biased?: Field study of the SAVRY and YLS/CMI in probation. *Psychological Assessment, 29*(6), 664.

Peters, C. S., Lampinen, J. M., & Malesky Jr, L. A. (2013). A trap for the unwary: Jury decision making in cases involving the entrapment defense. *Law and Human Behavior, 37*(1), 45.

Peters, M. J., Jelicic, M., & Merckelbach, H. (2006). When stereotypes backfire: Trying to suppress stereotypes produces false recollections of a crime. *Legal and Criminological Psychology, 11*(2), 327-336.

Petsko, C. D., & Bodenhausen, G. V. (2019). Race–Crime Congruency Effects Revisited: Do We Take Defendants' Sexual Orientation Into Account? *Social Psychological and Personality Science, 10*(1), 73-81.

Pfeifer, J. E., & Ogloff, J. R. (1991). Ambiguity and guilt determinations: A modern racism perspective 1. *Journal of Applied Social Psychology, 21*(21), 1713-1725.

Piatak, K. A. (2015). *Assessing Victim Blame: Intersections of Rape Victim Race, Gender, and Ethnicity.* (1591560 M.A.). East Tennessee State University, Ann Arbor. Retrieved from http://nottingham.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/assessing-victim-blame-intersections-rape-race/docview/1691128060/se-2?accountid=8018

Pica, E., Pettalia, J., & Pozzulo, J. (2017). The Influence of a Defendant's Chronological Age, Developmental Age, and Race on Mock Juror Decision Making. *Journal of Police and Criminal Psychology, 32*(1), 66-76. doi:http://dx.doi.org/10.1007/s11896-016-9201-1

Podsakoff, P., MacKenzie, S., & Lee, J. (1982). Podsakoff. *Psychological assessment, 17*, 222-232.

Qian, M. K., Quinn, P. C., Heyman, G. D., Pascalis, O., Fu, G., & Lee, K. (2019). A long-term effect of perceptual individuation training on reducing implicit racial bias in preschool children. *Child Development, 90*(3), e290-e305.

Quas, J. A., Bottoms, B. L., Haegerich, T. M., & Nysse-Carris, K. L. (2002). Effects of victim, defendant, and juror gender on decisions in child sexual assault cases 1. *Journal of Applied Social Psychology, 32*(10), 1993-2021.

Rachlinski, J. J., & Wistrich, A. J. (2017). Judging the judiciary by the numbers: Empirical research on judges. *Annual Review of Law and Social Science, 13*, 203-229.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 38*(1), 119-125.

Ritter, N. L. (2010). Understanding a widely misunderstood statistic: Cronbach's alpha. Paper presented at the annual meeting of the Southwest Education Research Association, New Orleans, February 18.

Sackeim, H. A., & Gur, R. C. (1978). Self-deception, self-confrontation, and consciousness. In G. E.

Schwartz & D. Shapiro (Eds.), *Consciousness and Self-regulation: Advances in Research* (Vol. 2, pp. 139-197). New York: Plenum Press.

Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: assessment and implications of generalized outcome expectancies. *Health Psychology, 4*(3), 219.

Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual-differences constructs. *Psychological Methods, 8*(2), 206.

Schoenberg, N. E., & Ravdal, H. (2000). Using vignettes in awareness and attitudinal research. *International Journal of Social Research Methodology, 3*(1), 63-74.

Schulz, K. F., & Grimes, D. A. (2002). Sample size slippages in randomised trials: exclusions and the lost and wayward. *The Lancet, 359*(9308), 781-785.

Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W. R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research, 58*(7), 935-943.

Shepherd, S. M., & Anthony, T. (2018). Popping the cultural bubble of violence risk assessment tools. *The Journal of Forensic Psychiatry & Psychology, 29*(2), 211-220.

Shiner, M., Carre, Z., Delsol, R., & Eastwood, N. (2018). The colour of injustice:'race', drugs and law enforcement in England and Wales.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Journal of Applied Psychology, 74, 107–120.

Silberkleit, G. (2021). Gendered race-based sexual stereotyping in mock juror decision making. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 82*(5-B), No-Specified. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc17&NEWS=N&AN=2020-97493-069

Sireci, S. G., Wainer, H., & Braun, H. (1998). Psychometrics. *ETS Research Report Series*, *1998*(1), i-56.

Skolnick, P., & Shaw, J. I. (1997). The OJ Simpson criminal trial verdict: Racism or status shield? *Journal of Social Issues, 53*(3), 503-516.

Skorinko, J. L., & Spellman, B. A. (2006). *Stereotypic crimes: How Group-Crime Associations Affect Memory and (sometimes) verdicts and sentencing.* Paper presented at the 1st Annual

Conference on Empirical Legal Studies Paper.

Smalarz, L., Madon, S., Yang, Y., Guyll, M., & Buck, S. (2016). The perfect match: Do criminal stereotypes bias forensic evidence analysis? *Law and Human Behavior, 40*(4), 420.

Smith, P., & Gudjonsson, G. H. (1995). Confabulation among forensic inpatients and its relationship with memory, suggestibility, compliance, anxiety, and self-esteem. *Personality and Individual Differences, 19*(4), 517-523.

Sommers, S. R. (2006). On racial diversity and group decision making: identifying multiple effects of racial composition on jury deliberations. *Journal of Personality and Social Psychology, 90*(4), 597.

Sommers, S. R., & Ellsworth, P. C. (2000). Race in the courtroom: Perceptions of guilt and dispositional attributions. *Personality and Social Psychology Bulletin, 26*(11), 1367-1379.

Sommers, S. R., & Ellsworth, P. C. (2001). WHITE JUROR BIAS: An Investigation of Prejudice Against Black Defendants in the American Courtroom. *Psychology, Public Policy, and Law, 7*(1), 201-229. doi:10.1037/1076-8971.7.1.201

Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 631.

Spector, P. E. (1992). *Summated rating scale construction quantitative applications in the social sciences*, Beverly Hills, CA: Sage Publications.

Stanton, J. M. (1998). An empirical assessment of data collection using the Internet. *Personnel Psychology, 51*(3), 709-725.

Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing continuous and dichotomous scoring of the Balanced Inventory of Desirable Responding. *Journal of Personality Assessment, 78*(2), 370-389.

Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology*.

Strosahl, K. D., Linehan, M. M., & Chiles, J. A. (1984). Will the real social desirability please stand up? Hopelessness, depression, social desirability, and the prediction of suicidal behavior. *Journal of Consulting and Clinical Psychology, 52*(3), 449.

Sunnafrank, M., & Fontes, N. E. (1983). General and crime related racial stereotypes and influence on juridic decisions. *Cornell Journal of Social Relations*.

Sweeney, L. T., & Haney, C. (1992). The influence of race on sentencing: A meta-analytic review of

experimental studies. *Behavioral Sciences & the Law, 10*(2), 179-195. doi:10.1002/bsl.2370100204

Tabachnick, B.G. and Fidell, L.S. (2007), *Using Multivariate Statistics* (5th ed.). New York: Allyn and Bacon.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology, 33*(1), 1-39.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago, IL: Nelson Hall

Tang, C. M., Nunez, N., & Estrada-Reynolds, V. (2020). Intellectual Disability Affects Case Judgment Differently Depending on Juvenile Race. *Journal of Police and Criminal Psychology, 35*(2), 228-239. doi:http://dx.doi.org/10.1007/s11896-019-09355-0

Taylor, D. M., & Jaggi, V. (1974). Ethnocentrism and causal attribution in a South Indian context. *Journal of cross-cultural Psychology, 5*(2), 162-171.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193.

Templeton, L. J., & Hartnagel, T. F. (2012). Causal attributions of crime and the public's sentencing goals. *Canadian Journal of Criminology and Criminal Justice, 54*(1), 45-65.

The Whitehouse (TWH) (2015). https://www.whitehouse.gov/blog/2015/07/15/president-obama-our-criminal-justice-system-isnt-smart-it-should-be

Thomas, C. (2008). Exposing the myths of jury service. *Criminal Law Review, 415*.

Thomas, C. (2010). *Are juries fair?* : Ministry of Justice London.

Toobin, J. (1996). The Marcia Clark verdict. *The New Yorker, 72*(26), 58-71.

Torres, S. (2009). Vignette methodology and culture-relevance: Lessons learned through a project on successful aging with Iranian immigrants to Sweden. *Journal of Cross-Cultural Gerontology, 24*(1), 93-114.

Ugwuegbu, D. C. E. (1979). Racial and evidential factors in juror attribution of legal responsibility. *Journal of Experimental Social Psychology, 15*(2), 133-146. doi:10.1016/0022-1031(79)90025-8

Uhrig, N. (2016). *Black, Asian and minority ethnic disproportionality in the criminal justice system in England and Wales* (Vol. 16): Ministry of Justice London.

UK Government (2021). Ethnicity Facts and Figures: Judges and non-legal members of the judiciary. Judges and non-legal members of the judiciary - GOV.UK Ethnicity facts and figures (ethnicity-facts-figures.service.gov.uk)

UK Government (2021). Ethnicity Facts and Figures: Police Officer Workforce. Police workforce - GOV.UK Ethnicity facts and figures (ethnicity-facts-figures.service.gov.uk)

UK Government (2021). Ethnicity Facts and Figures: Prison Officer Workforce. Prison officer workforce - GOV.UK Ethnicity facts and figures (ethnicity-facts-figures.service.gov.uk)

UK Government (2021). *Ethnicity Facts and Figures: Social workers for children and families.* Social workers for children and families - GOV.UK Ethnicity facts and figures (ethnicity-facts-figures.service.gov.uk)

Varelas, N., & Foley, L. A. (1998). Blacks' and Whites' Perceptions of Interracial and Intraracial Date Rape. *The Journal of Social Psychology, 138*(3), 392-400. doi:10.1080/00224549809600391

Vaske, J.J. (2008). Survey research and analysis: Applications in parks, recreation and human dimensions. *Venture Publishing Incorporated*, 23, 635.

Vispoel, W. P., & Kim, H. Y. (2014). Psychometric properties for the Balanced Inventory of Desirable Responding: Dichotomous versus polytomous conventional and IRT scoring. *Psychological assessment, 26*(3), 878.

Vispoel, W. P., & Tao, S. (2013). A generalizability analysis of score consistency for the Balanced Inventory of Desirable Responding. *Psychological assessment, 25*(1), 94.

Wiggins, J. S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instruction. *Journal of Consulting Psychology, 23*(5), 419.

Williams, W. (2020). Windrush lessons learned review. *Independent Review, March*, 104.

Wiseman, F. (1972). Methodological bias in public opinion surveys. *The Public Opinion Quarterly, 36*(1), 105-108.

Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication, 10*(3), JCMC1034.

Yalom, I. D., & Leszcz, M. (2020). The theory and practice of group psychotherapy. In: Basic books.

Yamamoto, S., & Maeder, E. M. (2017). Defendant and juror race in a necessity case: An ultimate attribution error. *Journal of Ethnicity in Criminal Justice, 15*(3), 270-284.

Ziegler, M., MacCann, C., & Roberts, R. (2012). *New perspectives on faking in personality assessment*: Oxford University Press.

Zinger, I. (2017). Annual report: Office of the correctional investigator 2016-2017. *Annual Report of the Office of the Correctional Investigator, 2016*.

Appendices

Appendix A

Percentage of Individuals Choosing a Black or White Criminal for Each Crime used in the study on 'General and Crime Related Racial Stereotypes and Influence on Juridic Decisions by Sunnafranks and Fonts (1983)

Crime	Ethnic Affiliation of Criminal	
	Black	White
Assault-Mugging	83	17
Fraud	35	65
Embezzlement	10	90
Grand-Theft Auto	81	19
Felony Vehicular Manslaughter	45	55
Assault on Police	72	28
Soliciting	95	5
Child Molestation	21	79
Rape	36	64
Counterfeiter	23	77

APPENDIX B

Search Strategies used to search each database.

Review Question: Do mock-jurors treat Black defendants and victims differently from White defendants and victims on decisions of guilt, blame and punishment? A systematic review

Database	Search String			
	Population	Intervention	Comparison	Outcome
1.PsychInfo	("mock jur*" or "simulated jur*" or "mock trial*" or "simulated trial*" or "jur* selection")	(Black* or "afro caribbean*" or caribbean* or african* or "afro american*" or "american african*" or "african american*" or BAME or BME or "ethnic* minorit*")	White*	(((race or racial* or ethnic*) adj4 (discriminat* or attitude* or bias*or prejudice*)) or (racial or racism or "rac* discriminat*" or "ethnic* discriminat*" or "rac* bias*" or "decision making" or "implicit bias" or "unconscious* bias*"))
2.Web of Science:	("mock jur*" or "simulated jur*" or "mock trial*" or "simulated trial*" or "jur* selection")	(Black* or "afro caribbean*" or caribbean* or african* or "afro american*" or "american african*" or "african american*" or BAME or BME or "ethnic* minorit*")	White*	(((race OR racial* OR ethnic*) NEAR4 (discriminat* OR attitude* OR bias*or prejudice*)) OR (racial OR racism OR "rac* discriminat*" OR "ethnic* discriminat*" OR "rac* bias*" OR "decision making" OR "implicit bias" OR "unconscious* bias*"))
3.Criminal Justice Database	("mock jur*" or "simulated jur*" or "mock trial*" or "simulated trial*" or "jur* selection")	(Black* or "afro caribbean*" or caribbean* or african* or "afro american*" or "american african*" or "african american*" or BAME or BME or "ethnic* minorit*")	White*	(((race OR racial* OR ethnic*) NEAR4 (discriminat* OR attitude* OR bias*or prejudice*)) OR (racial OR racism OR "rac* discriminat*" OR "ethnic* discriminat*" OR "rac* bias*" OR "decision making" OR "implicit bias" OR "unconscious* bias*"))
4.Psychology Database	("mock jur*" or "simulated jur*" or "mock	(Black* or "afro caribbean*" or caribbean* or african* or "afro american*" or	White*	(((race OR racial* OR ethnic*) NEAR4 (discriminat* OR attitude* OR bias*or prejudice*)) OR (racial OR racism OR "rac* discriminat*" OR "ethnic*

	trial*" or "simulated trial*" or "jur* selection")	"american african*" or "african american*" or BAME or BME or "ethnic* minorit*")		discriminat*" OR "rac* bias*" OR "decision making" OR "implicit bias" OR "unconscious* bias*"))
5.APA Articles	("mock jur*" or "simulated jur*" or "mock trial*" or "simulated trial*" or "jur* selection")	(Black* or "afro caribbean*" or caribbean* or african* or "afro american*" or "american african*" or "african american*" or BAME or BME or "ethnic* minorit*")	White*	(((race OR racial* OR ethnic*) NEAR4 (discriminat* OR attitude* OR bias*or prejudice*)) OR (racial OR racism OR "rac* discriminat*" OR "ethnic* discriminat*" OR "rac* bias*" OR "decision making" OR "implicit bias" OR "unconscious* bias*"))
6.Social Science Database	("mock jur*" or "simulated jur*" or "mock trial*" or "simulated trial*" or "jur* selection")	(Black* or "afro caribbean*" or caribbean* or african* or "afro american*" or "american african*" or "african american*" or BAME or BME or "ethnic* minorit*")	White*	(((race OR racial* OR ethnic*) NEAR4 (discriminat* OR attitude* OR bias*or prejudice*)) OR (racial OR racism OR "rac* discriminat*" OR "ethnic* discriminat*" OR "rac* bias*" OR "decision making" OR "implicit bias" OR "unconscious* bias*"))
7.International Bibliography of the Social Sciences (IBSS)	("mock jur*" or "simulated jur*" or "mock trial*" or "simulated trial*" or "jur* selection")	(Black* or "afro caribbean*" or caribbean* or african* or "afro american*" or "american african*" or "african american*" or BAME or BME or "ethnic* minorit*")	White*	(((race OR racial* OR ethnic*) NEAR4 (discriminat* OR attitude* OR bias*or prejudice*)) OR (racial OR racism OR "rac* discriminat*" OR "ethnic* discriminat*" OR "rac* bias*" OR "decision making" OR "implicit bias" OR "unconscious* bias*"))
8.ASSIA	("mock jur*" or "simulated jur*" or "mock trial*" or "simulated trial*" or "jur* selection")	(Black* or "afro caribbean*" or caribbean* or african* or "afro american*" or "american african*" or "african american*" or BAME or BME or "ethnic* minorit*")	White*	(((race OR racial* OR ethnic*) NEAR4 (discriminat* OR attitude* OR bias*or prejudice*)) OR (racial OR racism OR "rac* discriminat*" OR "ethnic* discriminat*" OR "rac* bias*" OR "decision making" OR "implicit bias" OR "unconscious* bias*"))
9. Criminology Collection	("mock jur*" or "simulated jur*" or "mock trial*" or	(Black* or "afro caribbean*" or caribbean* or african* or "afro american*" or "american african*" or	White*	(((race OR racial* OR ethnic*) NEAR4 (discriminat* OR attitude* OR bias*or prejudice*)) OR (racial OR racism OR "rac* discriminat*" OR "ethnic* discriminat*" OR "rac* bias*" OR "decision

	"simulated trial*" or "jur* selection")	"african american*" or BAME or BME or "ethnic* minorit*")		making" OR "implicit bias" OR "unconscious* bias*"))
10. Scopus	TITLE-ABS-KEY ("mock jur*" OR "simulated jur*" OR "mock trial*" OR "simulated trial*" OR "jur* selection") AND	(Black* OR "afro caribbean*" OR caribbean* OR african* OR "afro american*" OR "american african*" OR "african american*" OR bame OR bme OR "ethnic* minorit*") AND	White* AND	(race OR racial* OR ethnic* PRE/4 discriminat* OR attitude* OR bias*or AND prejudice*) OR (racial OR racism OR "rac* discriminat*" OR "ethnic* discriminat*" OR "rac* bias*" OR "decision making" OR "implicit bias" OR "unconscious* bias*") AND (LIMIT-TO DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "re")) AND (LIMIT-TO (SUBJAREA, "PSYC") OR LIMIT-TO (SUBJAREA, "SOCI") OR LIMIT-TO (SUBJAREA, "ARTS") OR LIMIT-TO (SUBJAREA, "DECI")) AND (LIMIT-TO (LANGUAGE "English"))
11. ProQuest Dissertations & Theses A&I	("mock jur*" or "simulated jur*" or "mock trial*" or "simulated trial*" or "jur* selection")	(Black* or "afro caribbean*" or caribbean* or african* or "afro american*" or "american african*" or "african american*" or BAME or BME or "ethnic* minorit*")	White*	(((race OR racial* OR ethnic*) NEAR4 (discriminat* OR attitude* OR bias*or prejudice*)) OR (racial OR racism OR "rac* discriminat*" OR "ethnic* discriminat*" OR "rac* bias*" OR "decision making" OR "implicit bias" OR "unconscious* bias*"))

Appendix C

Data extraction form used in systematic review

Author	Year	Country	Population (Include Sample size; gender; Race Profession).	Offence	Scenario Setting (real/fictious)	Presentation of Scenario	Race manipulation	Study design	Outcome Measure	Findings

Appendix D

Domains of risk of bias

Study	Did the study address a clearly focused question / issue?	Is the research method (study design) appropriate for answering the research question?	Is the method of selection of the subjects clearly described?	Could the way the sample was obtained introduce (selection) bias?	Was the sample of subjects' representative with regard to the population to which the findings will be referred?	Was the sample size based on pre-study considerations of statistical power?	Was a satisfactory response rate achieved?	Are the measurements (questionnaires) likely to be valid and reliable?	Was the statistical significance assessed?	Are confidence intervals given for the main results?	Could there be confounding factors that haven't been accounted for?	Can the results be applied to your organization?
1.Gamblin & Kehn (2021)	Y	Y	Y	Y	Y	Y	U	U	Y	N	Y	N
2. Silberkleit (2021)	Y	Y	Y	Y	U	N	U	U	Y	N	Y	N
3. Ellis (2020)	Y	Y	Y	Y	Y	U	U	U	Y	Y	Y	N
4. Maeder, & McManus (2020)	Y	Y	Y	Y	Y	U	U	U	Y	N	Y	N
5. Maeder, Yamamoto & McLaughlin (2020)	Y	Y	Y	Y	Y	Y	U	Y	Y	N	Y	N
6.Tang et al. (2020)	Y	Y	Y	Y	U	U	U	U	Y	Y	Y	N
7. Bell (2019)	Y	Y	Y	Y	U	U	N	U	Y	N	Y	N
8. Burke (2019)	Y	Y	Y	Y	Y	U	U	U	Y	N	Y	N

9. Call (2019)	Y	Y	Y	Y	U	Y	U	U	Y	Y	Y	N
10. Kurinec & Weaver, III (2019)	Y	Y	Y	Y	U	U	U	U	Y	N	Y	N
11. Maeder & Yamamoto (2019a)	Y	Y	Y	Y	Y	U	U	U	Y	N	Y	N
12. Maeder & Yamamoto (2019b)	Y	Y	Y	Y	U	U	U	U	Y	N	Y	N
13. Nesbitt (2019)	Y	Y	Y	Y	Y	U	U	U	Y	N	Y	N
14. Maeder & Ewanation (2018)	Y	Y	Y	Y	Y	U	U	U	Y	N	Y	N
15. McManus, Maeder & Yamamoto (2018)	Y	Y	Y	Y	Y	U	U	U	Y	Y	Y	N
16. Armenta (2017)	Y	Y	Y	Y	N	U	U	U	Y	N	Y	N
17. DeCesare (2017)	Y	Y	N	Y	U	U	U	U	Y	N	Y	N
18. Leippe, et al. (2017)	Y	Y	Y	Y	U	U	U	U	Y	N	Y	N

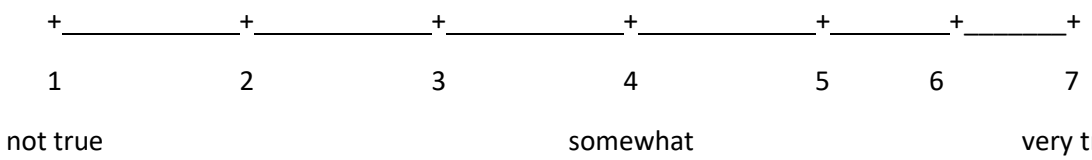
19. Pica, et al. (2017)	Y	Y	Y	Y	Y	U	U	U	Y	Y	Y	N
20. Yamamoto, & Maeder (2017)	Y	Y	Y	Y	Y	U	U	U	Y	Y	Y	N
21. David (2016)	Y	Y	Y	Y	U	U	U	U	Y	N	Y	N
22. Maeder, et al. (2016)	Y	Y	Y	Y	Y	U	U	Y	Y	N	Y	N
23. Dodson, (2015)	Y	Y	Y	Y	Y	U	U	U	Y	N	Y	N
24. Glaser et al. (2015)	Y	Y	Y	U	U	U	U	U	Y	Y	Y	N
25. Kluwe (2015)	Y	Y	Y	Y	U	Y	U	U	Y	N	Y	N
26. Maeder, Yamamoto, & McManus (2015)	Y	Y	Y	Y	Y	U	U	U	Y	N	Y	
27. Maeder, Yamamoto & Saliba (2015)	Y	Y	Y	Y	U	U	U	U	Y	N	Y	N
28. Piatak (2015)	Y	Y	Y	Y	U	U	U	Y	Y	N	Y	N

Key: Y = Yes; N = No; U = Can't tell.

Appendix E

BIDR Version 16 - Short Form

Using the scale below as a guide, write a number beside each statement to indicate how true it is.



____ 1. I have not always been honest with myself.

____ 2. I always know why I like things.

____ 3. It's hard for me to shut off a disturbing thought.

____ 4. I never regret my decisions.

____ 5. I sometimes lose out on things because I can't make up my mind soon enough.

____ 6. I am a completely rational person.

____ 7. I am very confident of my judgments

____ 8. I have sometimes doubted my ability as a lover.

____ 9. I sometimes tell lies if I have to.

____ 10. I never cover up my mistakes.

____ 11. There have been occasions when I have taken advantage of someone.

____ 12. I sometimes try to get even rather than forgive and forget.

____ 13. I have said something bad about a friend behind his/her back.

____ 14. When I hear people talking privately, I avoid listening.

____ 15. I never take things that don't belong to me.

____ 16. I don't gossip about other people's business.

Appendix F

University of Nottingham
UK | CHINA | MALAYSIA

Email: FMHS-ResearchEthics@nottingham.ac.uk

Faculty of Medicine & Health Sciences
Research Ethics Committee
c/o Faculty PVC Office
School of Medicine Education Centre
B Floor, Medical School
Queen's Medical Centre Campus
Nottingham University Hospitals
Nottingham, NG7 2UH

27 April 2018

Donna Spencer
Doctorate in Forensic Psychology Student
c/o Dr Nigel Hunt
Room B19 YANG Fujia Building
Jubilee Campus
Wollaton Road
Nottingham
NG8 1NN

Dear Ms Spencer

Ethics Reference No: 289-1803 – **please always quote**	
Study Title: The effect of the race of defendant and victim on criminal justice workers' judgement of violent assault crime.	
Chief Investigator/Supervisor: Dr Nigel Hunt, Associate Professor, Dr Vince Egan Associate Professor, Forensic Psychology Practice, Centre for Famlily and Forensic Psychology. Division of Psychiatry and Applied Psychology, School of Medicine.	
Lead Investigators/student: Donna Spencer, Top-up Doctorate in Forensic Psychology, 2nd Year	
Type of Study: PhD, internet mediated	
Proposed Start Date: 01/06/2018	**Proposed End Date:** 01/09/2018 3 mths
No of Subjects: 90	**Age:** 18+years

Thank you for submitting this application and the following documents were received:

- FMHS REC Application form and supporting documents version 1.0: 25.04.2018

These have been reviewed and are satisfactory and the study has been given a favourable opinion.

A favourable opinion has been given on the understanding that:

1. The protocol agreed is followed and the Committee is informed of any changes using a notice of amendment form (please request a form).
2. The Chair is informed of any serious or unexpected event.
3. An End of Project Progress Report is completed and returned when the study has finished (Please request a form).

Yours sincerely



Professor Ravi Mahajan
Chair, Faculty of Medicine & Health Sciences Research Ethics Committee

Appendix G

Instructions to the mock-juror (criminal justice professionals)

On the following pages are a condensed case description of a criminal case. The study requires you to imagine that you are part of this jury. A jury consists of twelve people (jurors) who listen to a case and decide whether the person(s) on trial are innocent or guilty of the crime they are accused of committing. Your task is as follows:

1. Complete the demographic questionnaire.
2. Complete the self-assessment questionnaire.
3. Read the condensed case description.
4. Complete the two questionnaires about the case.

This task should take approximately 20-30 minutes to complete.

- Note answering all questions will help in my later statistical analysis.

Appendix H

Juror Pre-Trial Demographic Questionnaire:

1. What is your Age?

2. What is your Sex (please select appropriate box)?
   a. ☐Male b. ☐Female c. ☐Other Please specify _____

3. What is your Ethnicity?
   a. ☐White b. Black ☐ C. ☐Asian
   d. ☐ Other. Please specify _____

4. Which service do you work for?
   a. Probation Service b. Youth Offending Service c. Children's Social Care
   b. Other(specify)
5. What is your role?
   a. Probation Officer ☐ b. Youth Offending Worker ☐
   c. Social Worker ☐ d. Other (specify): _____

Appendix I

Court Case Vignettes to be read by participants.

Court Case 1:

The attack happened at around 3.45pm on 20th January 2015. Mr Taylor, a Black/White/ Asian man in his mid-20s from Marks Estate, was travelling on bus number 24 on his way to meet a friend. As Mr Taylor got off the bus he was attacked by a man in his late teens to mid-twenties. The man was wearing a hoodie, baseball cap and bandana. CCTV footage showed that Mr Taylor was repeatedly punched, kicked, and stamped on, before he got away. Mr Taylor sustained injuries to his head and torso, which required hospital treatment. On 28th June 2015, Mr Smith (also known by his street name 'Razor'), a Black/White/ Asian man aged 24 from the Cleveland Estate, was arrested and charged with Grievous Bodily Harm (GBH). Mr Smith pleaded not guilty.

Prosecution's summary:

Mr Taylor sustained head injuries and bruising to his torso. Mr Smith (also known as 'Razor') was picked out from an identity parade; is the same height as the attacker; has known affiliations with the Cleveland Park Bruisers; has inconsistent accounts of where he was on the afternoon of the offence; and had a previous altercation with the victim, Mr Taylor.

Defence's Summary:

The evidence presented by the prosecution against Mr Smith is inconclusive. Mr Smith's cousin testified under oath that Mr Smith was with him on the afternoon of the offence. Mr Smith lives within the locality of the Cleveland Park Bruisers, and it is therefore almost impossible for him not to bump into someone from the Bruisers in his daily interactions. Mr Smith denies being involved in the attack on Mr Taylor.

Appendix J

Post-Trial Questionnaire:

Instructions: Please read each question carefully and check the box that corresponds to your answer.

1. Based on what you have read and the instructions that you have been provided, how likely do you find the defendant guilty of actual bodily harm?

1 2 3 4 5 6 7

Key:

1=not at all guilty 7=definitely guilty

2. How confident are you in your verdict?

Not at all slightly moderately extremely totally

3. What most influenced your verdict?

☐Prosecution evidence?

☐Defence evidence?

☐Other evidence? Please specify

4. How much blame do you think the victim should take for the offence?

None a little half a lot all

5. To what extent do you believe the defendant's behaviour was due to situational factors (e.g., stress), or internal factors (e.g., personality characteristics)?

1 2 3 4 5 6 7

Key:

1=totally situational; 2=extremely situational; 3=generally situational; 4=partly situational/ partly internal; 5= generally internal; 6=extremely internal; 7=total internal

6.To what extent do you think that their criminal behaviour was due to factors beyond their control?

Not at all slightly moderately extremely totally

7. What sentence do you think he should be given?

☐None ☐Prison ☐Community Sentence ☐Other. Please specify _____

8.If you believe that the defendant should receive a prison sentence how long do you think this should be?

a). _____Years or _____Months or _____Days

b). If you believe that the defendant should receive a community sentence how long do you think this should be?

_____Years or _____Months or _____Days

Other please specify type of sentence and how many months or years this should be?

c). _____Years or _____Months or _____Days

If the defendant is found guilty, how typical do you think the offender do you think the defendant is for this sort of crime?

Not at all slightly moderately extremely totally

9. Do you think that the defendant should be convicted?

☐Yes

☐No

10. What are the reasons behind your decisions?

Appendix K

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

PARTICIPANT INFORMATION

Division of Psychiatry & Applied Psychology

School of Medicine, Faculty of Medicine & Health Sciences

Project Title: The effect of the race of defendant and victim on criminal justice workers' judgements of violent assault crime

Researcher: Donna Spencer (*donna.spencer@nottingham.ac.uk)*

Supervisor: Professor Nigel Hunt (nigel.hunt@nottingham.ac.uk).

Ethics Reference Number: … *[to be inserted following ethical review]*

This is an invitation to take part in a research study examining decision-making in criminal justice professionals (probation officers/youth offending case manager/ social worker/ other) using a mock-juror method. This information sheet tells you about the research study and explains why the study is being done and why you are being invited to participate. It also tells you what will happen if you agree to participate in this study.

What is the project about?

This research is concerned with looking at the mock-juror decisions of criminal justice professionals. Whilst some research has already been done within the area of mock-juror decisions, little work has been done on the decision making of criminal justice professionals. Therefore, the mock-juror decisions of probation officers/ youth offending case manager/ social worker may be very different from other groups. Also, much of the work that has been done was done outside of the UK;

therefore, it is possible that there are cultural differences between criminal justice professionals within England and Wales criminal justice professionals from other countries. Lastly, as research with e.g., probation officers as mock jurors was done many years ago, the decisions of this professional group may have changed with the passage of time. The findings from this research aims to add clarity to mock-juror research and shed further light on the mock-juror decision making of criminal justice professionals/ professional involved in criminal justice work within the UK.

Who is being asked to take part, and why?

Probation officers working within the National Probation Service England; youth offending service case managers; social workers/ and other criminal justice professionals are being asked take part in this research, to assess their mock-juror views and judgements on a violent offence.

Do I have to take part?

Your participation in the research is completely voluntary, and you should not feel obliged to participate. If you do decide to take part, you have the right to withdraw during or after the study should you choose to do so. All data will be anonymized.

What will I be asked to do?

If you decide to participate in this research, you will be required to take on the role of mock-juror. A mock juror is someone who mimics the role of an actual juror member, by considering the evidence presented in a court trial, and following the evidence presented, decides on the innocence or guilt of the defendant(s). If you decide to participate in the current research, first you will be asked to complete a demographic questionnaire like what is used to pool actual jurors. Following the completion of this demographic questionnaire, you will be asked to take on the role of a juror and read a short summary of a criminal case, and then complete a questionnaire about the case you read. The questions will explore your verdict and perceptions surrounding the case. Your participation will take approximately 20-30 minutes. Following participation, on a later date within a few months you may be asked to participate in a further interview via email. All aspects of participation are entirely voluntary.

Will the research be of any personal benefit to me?

The findings from this research may help shed further light on probation officer's/ youth offending case managers/ social worker's mock-juror decisions.

What will happen to the information I provide?

Once your questionnaire is submitted it will be anonymised in line with Data Protection Legislation. This will maintain confidentiality and prevent you (and others) from being directly or indirectly identified through e.g., your work location or demographic information. Whilst we will take steps to ensure that your online questionnaire is gathered and retained securely using a secure website, which only I and my supervisor have access to, it remains a possibility that these security measures can be breached by hackers. A further limitation of Internet based research is that your anonymity may be compromised if you complete the questionnaire within a public space in view of others. For example, if you decide to use a computer within an organisation (e.g., work) that monitors computer use, your responses may not be entirely confidential. Consequently, you may wish to take additional

measures to maintain confidentiality whilst in the process of completing the questionnaire. Note that the responses from your questionnaire along with that of others will later be uploaded from the secure website and analysed using the Statistical Package for Social Sciences (SPSS). Access to this statistical database will be password protected and accessible only to me and my supervisor (Nigel Hunt). At the end of the project your raw data alongside that of other people involved in the research will be retained securely (password protected) and electronically by the University of Nottingham for 7 years before being destroyed. SPSS). Access to this statistical database will be password protected and accessible only to me and my supervisor (Nigel Hunt). At the end of the project your raw data alongside that of other people involved in the research will be retained securely (password protected) and electronically by the University of Nottingham for 7 years before being destroyed).

What will you do with the data?

The results from this statistical analysis will be presented as a written report and submitted to the University as an assignment for my professional Doctorate qualification in Forensic Psychology. The findings from this research may be published in an academic journal.

If you have any questions or concerns that you want to raise regarding the research either before or after your participation, you can do so by contacting myself Donna Spencer, or Dr Nigel Hunt, on the email addresses provided.

THANK YOU FOR YOUR PARTICIPATION

If you have any queries or complaints about this study, please contact my supervisor in the first instance on nigel.hunt@nottingham.ac.uk. If this does not resolve the query to your satisfaction, please write to:

Administrative Support

Faculty of Medicine & Health Sciences Research Ethics Committee

c/o Faculty PVC Office

B Floor, Medical School (nr Bridge)

QMC Campus, Nottingham University Hospitals

Nottingham

NG7 2UH

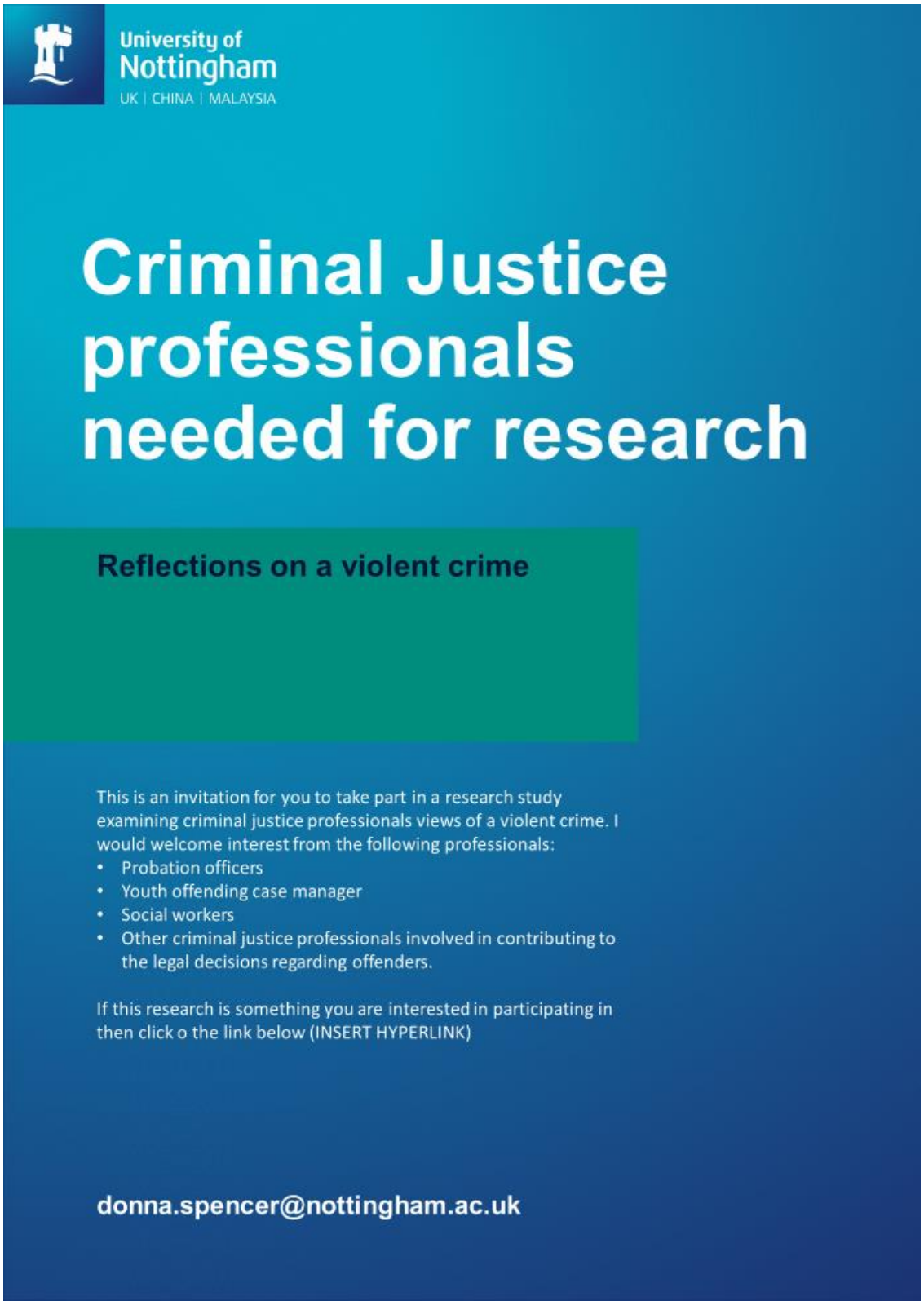
e-mail: louise.sabir@nottingham.ac.uk

We do not believe that there are foreseen risks in participating in this research. However, if you do become upset because of participating in this research, you are advised to seek support from your GP.

It is anticipated that the results from this research will be available from October 2019. You will be made able to request information about the research by emailing either me at donna.spencer@nottingham.ac.uk; or nigel.hunt@nottingham.ac.uk.

Appendix L

Recruitment poster



Appendix M

Nine different conditions used in study

Condition	Defendant Race	Victim Race
1	Black	Black
2	Black	White
3	Black	Asian
4	White	Black
5	White	White
6	White	Asian
7	Asian	Black
8	Asian	White
9	Asian	Asian

Appendix N

PARTICIPANT CONSENT

University of Nottingham

An Examination of Criminal Justice Workers' Mock Juror Decisions

Researcher: Donna Spencer

PARTICIPANT CONSENT FORM

Please answer every question with a YES or NO by selecting the appropriate box to create a cross.

	Please complete the following:	**YES**	**NO**
1	I have read the information sheet that describes this research.	☐	☐
2	I have been given the opportunity to ask questions and discuss this research.	☐	☐
3	I have received satisfactory answers to all my questions.	☐	☐
4	I have received sufficient information about this research.	☐	☐
5	I understand that I do not have to take part in this research.	☐	☐
6	I understand that I am free to withdraw from the research without giving a reason, up until I click submit.	☐	☐
7	I understand that after I have clicked to submit my questionnaire, it becomes anonymized and can no longer be withdraw as it cannot be traced.	☐	☐
8	I agree that my questionnaire will be collected stored securely online described in the information sheet.	☐	☐

9	I agree for the information gathered from my questionnaire to be used for statistical analysis using the Statistical Package for Social Sciences (SPSS).	☐	☐
10	I understand that some documents from the study may be looked at by responsible people appointed by The University of Nottingham, who must make sure (as Research Governance sponsor) that the research is being run properly. I give permission for this group to have access to the necessary information.	☐	☐
11	I agree that data from my questionnaire, including quotes that have been anonymised to uphold my anonymity, can be used in the research report and included in publications.	☐	☐
12	I understand that the personal information I provide will only be used for the purposes of this project and not transferred to an organisation outside of the University of Nottingham. The information will be treated confidentially and handled in accordance with the provisions of the Data Protection Act 1998.	☐	☐
13	I agree that my data from this study can be shared with other researchers in the future provided that my anonymity is protected	☐	☐
14	Do you agree to take part in this study research?	☐	☐

Appendix O

Debrief Sheet

This research is designed to examine the effect of the race of defendant and victim on criminal justice professionals' judgements of a violent assault crime. Previous research within the field of mock-juror decision making has shown that extra-legal factors, such as race, can influence juror/jury decisions (Green, 1964). Research conducted within the United States has shown that when primed, Probation Officers showed racial bias in their decision-making (Graham & Lowery, 2004) such that White witnesses tended to be perceived as more credible (Johnson, 1996). The present study is concerned with examining whether this bias exists amongst criminal justice professionals in the UK. The effect of the race of the defendant (Black/White/Asian) on criminal justice professionals views of a violent crime were assessed in the current study. The race of criminal justice professionals was controlled. It is hypothesized that the race of the defendant will influence criminal justice professionals' decisions on the defendant's guilt; culpability and perceived seriousness of the offence. Once the current study has completed data collection, the views of all the criminal justice professionals who participated in this research will be collated and statistically analysed to see whether there is a difference in their judgements based on the race of defendant and victim. This research area is important as Criminal justice professionals are involved in the sentence planning; parole and release of offenders.

If after reading this debrief you no longer want to participate in this research, you have the right to withdraw and can do so by notifying me by email on donna.spencer@nottingham.ac.uk. If you choose to withdraw, your data will be destroyed and not included in the final analysis. There is no negative consequence if you decide to withdrawal from this study. Please remember that all information within this research study is anonymised to maintain the confidentiality of participants.

If you have any further questions or queries regarding this research study, then please feel free to contact me at the email address provided.

I would like to thank you for taking the time to participate in this research.

Donna Spencer
DForen Psychology Student
Division of Psychiatry and Applied Psychology
School of Medicine
University of Nottingham
Email address: donna.spencer@nottingham.ac.uk

Supervisor: Dr Nigel Hunt CPsychol AFBPsS FRHistS
Associate Professor
Division of Psychiatry and Applied Psychology
School of Medicine
University of Nottingham
Email address: nigel.hunt@nottingham.ac.uk

Appendix P

Perceived guilt and attributed blame by defendant's race, victim's race, and profession

Source	Dependent Variable	*F*	*df*	*p* value	η_p
Corrected Model	Guilt	.48	8, 399	.87	.01
	Blame	.86	8, 399	.55	.02
Intercept	Guilt	1888.88	1, 390	<.001	.86
	Blame	1162.27	1, 390	<.001	.79
Defendant race	Guilt	.88	2, 390	.41	.01
	Blame	1.19	2, 390	.30	.01
Victim race	Guilt	.10	2, 390	.90	.001
	Blame	1.58	2, 390	.21	.01
Defendant race * victim race	Guilt	.49	4, 390	.74	.01
	Blame	.31	4, 390	.87	.004

Source	Dependent Variable	*F*	df	*p* value	η_p
Corrected Model	Guilt	1.91	14, 398	.024	.06
	Blame	.91	14, 398	.54	.03
Intercept	Guilt	1439.4	1, 384	<.001	.79
	Blame	877.00	1, 384	<.001	.69
Profession	Guilt	3.53	4, 384	.008	.03
	Blame	1.75	4, 384	.14	.02
defendant's race	Guilt	.72	2, 384	.49	.004
	Blame	.448	2, 384	.65	.002
Profession * defendant's race	Guilt	1.33	8, 384	.23	.03
	Blame	.63	8, 384	.75	.02

Source	Dependent Variable	*F*	df	*p* value	η_p
Corrected Model	Guilt	1.64	14, 398	.06	.06
	Blame	1.78	14, 398	.040	.06
Intercept	Guilt	1474.22	1, 384	<.001	.79
	Blame	936.91	1, 384	<.001	.71
Profession	Guilt	3.93	4, 384	.004	.04
	Blame	2.09	4, 384	.08	.02
Victim's race	Guilt	2.08	2, 384	.13	.01
	Blame	.81	2, 384	.45	.004
Profession * victim's race	Guilt	.76	8, 384	.64	.02
	Blame	1.96	8, 384	.050	.04

df: degrees of freedom

Appendix Q

Perceived guilt and attributed blame by defendant's race, victim's race among White CJWs

Source	Dependent Variable	*F*	*df*	*p* value	η_p
Corrected Model	Guilt	.48	8, 325	.869	.01
	Blame	.85	8, 325	.55	.02
Intercept	Guilt	1888.88	1, 317	<.001	.86
	Blame	1162.27	1, 317	<.001	.79
Defendant race	Guilt	.88	2, 317	.41	.01
	Blame	1.19	2, 317	.30	.01
Victim race	Guilt	.10	2, 317	.90	.001
	Blame	1.58	2, 317	.21	.01
Defendant race * victim race	Guilt	.49	4, 317	.74	.01
	Blame	.31	4, 317	.87	.004

Appendix R

Memory Questionnaire

The following questions will test your memory of the case that you have just read.

1. What was the offence?
2. How did the offence take place?
3. Were there others involved in the offence?
4. If the offender had a weapon, what weapon did they have?
5. The victim was ended up in a coma?
    a. ☐True ☐False
6. The victim was carrying a weapon
    a. ☐True ☐False
7. How old was the defendant? ____
8. The victim was female
    a. ☐True ☐False
9. What was the sex of the defendant?
    a. ☐Male
    b. ☐Female
10. The victim stabbed?
    a. ☐True ☐False
11. What was the race of the defendant?
12. What items were found on the defendant?
13. How old was the victim?
14. What was the sex of the victim?
    a. ☐Male
    b. ☐Female
15. The victim was a pensioner?
    a. ☐True ☐False
16. What was the race of victim?
17. If the victim sustained injuries, what injuries did the victim sustain in the attack?
18. What evidence did the prosecutor state points to the defendant's guilt?
19. What evidence did the defence state points to the defendant's innocence?

Appendix S

Perceived guilt and attributed blame among White CJWs

	Dependent Variable	*F*	df	*p* value	η_p
Corrected Model	Guilt	3.79	4, 325	.005	.04
	Blame	1.60	4. 325	.17	.02
Intercept	Guilt	1529.31	1, 321	<.001	.79
	Blame	932.84	1, 321	<.001	.70
Profession	Guilt	3.79	4, 321	.005	.04
	Blame	1.61	4, 321	.17	.02

df: degrees of freedom

Appendix T

Broad themes extracted from the answers provided by CJWs who chose other type of sentence for the described defendant

Insufficient evidence	Background information on defendant /situational factor needed	Custodial sentence following background check	Suspended sentence order (with/without prison)
"not convinced of guilt based on the information provided" "I would want far more than is provided (pre-cons, victim impact, greater offence analysis, remorse" "Insufficient proof …. so too early to answer" "Bail pending the prosecution obtaining further evidence" "Very difficult to sentence without knowing the lasting impact of the offence" "Defendants whereabout at the time are unconfirmed" "On the face of it this would most likely result in a prison sentence, but it is possible that meaningful work could be carried out with a community sentence"	"pre-sentencing report needs to be done … to understand what the most effective option for rehabilitation is" "Without knowing what previous convictions, mitigation or pre-sentence reports this is not an answer I can give" "If not a first-time offender consider custody, … If first time offender, consider alternative to custody such as a SSO" "Depends on other factors such as previous offending, Gravity score" "Has ... defendant got any mental health conditions, learning difficulties that to me would reflect on the outcome of the verdict" "Drug and alcohol issues would also need to be considered" "I need more information about …previous offences and engagement with probation and/or youth offending" "For GBH it should be prison but there needs to be a focus on why the act was committed	"If … offender did commit assault for no reason then a custodial sentence" "If found guilty …, then he should face a custodial sentence" "If it is deemed, he/she is a danger to the community, a custodial sentence may be offered"	"GBH should get a minimum SSO because it's a serious crime" "Possibly a Suspended Sentence Order but I don't know his previous" "1-year custodial sentence suspended for 2 years" "SSO as the offence passes the threshold of Risk of Serious Harm as he possesses a medium risk to the general public" "A SSO is also deemed appropriate as it gives the person a chance to grow and change but also presents the alternative of being sent to custody to demonstrate the seriousness of the crime" "Also, he should be issued with 20 Rehabilitation Activity Requirements…"

Compensation (with community sentence)	Punishment combined with rehabilitation or restorative justice	Possibly not guilty	Beyond remit
“Without knowledge of his previous history, … he could have compensation Order, period of HDC” “Supervised community sentence and made to pay comp to victim”	“Depending on previous convictions and sentencing guidelines, an order combined with punishment and rehabilitation” “If guilty the sentence should focus on rehabilitative/risk reduction work.” “Whether community or prison, there should be an element of restorative justice involved” “Get him to apologise to the victim and explain why he did it. I don't believe in the prison system”	“As yet in the evidence there is no proof of guilt beyond a reasonable doubt.” “I’m not convinced the defendant is guilty” “Not sure if he was guilty based on the evidence supplied” “I believe the suspect cannot be found guilty, thus this question makes no sense”	“I’m not in a position to judge or assume who goes to prison or who don’t” “It is not the responsibility of juries to pass sentences”