# What's in the Box? Punishment and Insanity in the Canadian Jury Deliberation Room

by

Susan Yamamoto

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Psychology

Carleton University

Ottawa, Ontario

**Abstract**

There is a longstanding culture of hostility toward the insanity defence in Canada, where jurors may be hesitant to find defendants Not Criminally Responsible on Account of Mental Disorder (NCRMD) in a criminal trial. While this hesitancy is in part attributable to misinformation about the defence, it might also be seated in moral intuitions about fair punishment. Researchers have made gains in understanding individual juror decision-making in insanity cases, but the jury deliberation room remains a black box. In this mixed-methods study, Canadian jury eligible participants ($N = 83$) completed attitudinal measures, read a fictional murder case involving a claim of NCRMD, then took part in 45-minute deliberation sessions. Study 1 examined the relationship between punishment orientation, insanity defence attitudes, and individual verdict decisions, and addressed the utility of a hypothetical self-exclusion question (i.e., Challenge for Cause). The Challenge for Cause question was not a useful gauge of individual verdict decisions. However, punishment prone orientations were correlated with higher confidence in a guilty verdict. In Study 2 two independent coders were trained to assess each individual utterance for the presence or absence of categories listed in a codebook that was created a-priori, which yielded fair inter-rater reliability. Hierarchical linear modeling analyses showed that punishment-prone mock jurors were less likely to defer to the authority of the psychiatrist testifying for the Defence. Those who were more punishment-prone also had a lower frequency of Defence position-taking utterances. Finally, in Study 3, a qualitative description of key-word flagged utterances (using the Moral Foundations Dictionary) demonstrated that mock jurors relied on moral intuitions about authority, harm, and fairness in justifying their positions. Study 3 also revealed five general

categories with respect to what would happen to the defendant after the trial: the effectiveness of prisons, the conditions of prisons, the jury's duty in considering punishment, desires for rehabilitation, and desires for incapacitation. Overall, findings imply that mock jurors' decisions stem partially from moral conceptualizations of insanity rather than from evidence alone.

**Acknowledgments**

Words cannot express my gratitude toward my mentor, Dr. Evelyn Maeder, whose contagious passion and dedication to her students is inspirational. For seven years, her fierce work ethic, talent, and compassion have energized me along this turbulent road. It has truly been a joy and privilege to be part of the Legal Decision-Making Lab.

I am grateful for the time and efforts of my committee members Dr. Craig Bennell and Dr. Kevin Nunes in pushing me to strengthen this project. Additionally, I would like to thank Dr. Dale Spencer and Dr. Mona Lynch for their time and insights. I have been fortunate to know many amazing researchers in the Psychology Department at Carleton; in particular I thank Dr. Andrea Howard and Dr. Deanna Whelan for sharing their priceless statistical know-how. I would be remiss not to express appreciation to Etelle Bourassa, Marilyn Ginder, and Robin Dunbar for their guidance in navigating university life over the years.

Finally, thank you to my friends and family for their unconditional support through this long haul. I am eternally grateful to the Merrimans and the Milners for looking out for me, and especially to my parents, sisters, and husband whose care and patience have been sustaining. I will never forget the kindness of my peers, Laura McManus, Kendra McLaughlin, Logan Ewanation, and Natasha Korva. A special thanks to Femi Carrington, Traleena Rouleau, Kevin Simas, and the many other students who contributed to this project. As they say in Japan, Otsukaresama desu – Thank you all for your hard work.

## Table of Contents

## List of Tables

## List of Appendices

## Introduction

"Our collective conscience does not allow punishment where it cannot impose blame."

- Judge David Bazelon

Criminal law dictates that some acts are so harmful as to constitute crimes against society and therefore must be punished. A crime has two main components: the actus reus (the physical act) and mens rea (guilty mind). In the case of murder, it must be shown that the defendant voluntarily caused the death of the victim. If one was not in physical control of his or her body, for instance due to reflex, then it might be said that there is no actus reus. Of equal significance, a person must *wilfully* or intentionally have caused the victim's death. Consequently, the Canadian legal system takes a clear stance on the issue of criminal culpability and mental disorder. According to Section 16 of the Criminal Code: "No person is criminally responsible for an act committed…while suffering from a mental disorder that rendered the person incapable of appreciating the nature and quality of the act or omission, or of knowing it was wrong" (Criminal Code of Canada, 1985). Hence a person may be found Not Criminally Responsible on Account of Mental Disorder (NCRMD) if the party raising the issue can prove it is more likely than not that, during the crime, the defendant had a mental disorder that precluded a guilty mind. Rather than traditional punishment via the criminal justice system, a successful NCRMD claim will result in psychiatric care or in some cases absolute discharge.

Incidentally, this provision is in tension with a longstanding culture of hostility toward the insanity defence in Canada (Maeder, Yamamoto, & Fenwick, 2015) and the United States (Hans, 1986). Research confirms that there are widespread negative insanity defence attitudes, which are at least partially rooted in legal misconceptions

(Bloechl, Vitacco, Neumann, & Erickson, 2007; Skeem, Louden, & Evans, 2004) and stereotypes associating mental disorder with violence (Silver, Cirincione, & Steadman, 1994). Despite the rarity of NCRMD (i.e., up to 6.08 per 1000 among decisions averaged over five years in three provinces; Crocker et al., 2015) and the strict review process, the public tends to see it as a frequently exploited loophole that sets dangerous offenders free (Skeem et al., 2004). Unfortunately, this prejudice can result in mock jurors' inability to correctly apply this law when appropriate (Bloechl et al., 2007; Maeder et al., 2015). Because the law is unequivocal that there can be no guilt without mens rea, we should consider at least three potential courses of action.

First, we could consider the public's rejection of the insanity defence as substantive information about the appropriateness of current legal standards. In theory, a fundamental function of the jury system is to bring community sentiment into the courtroom, constituting participation in democracy. Interestingly, in coming to a decision, juries are not obligated to follow the law, a little-known practice called 'nullification'. While nullification can function to protect a defendant, juries do sometimes seek greater punishment than is recommended by the Judge (e.g., *Liebeck v. McDonald's Restaurants,* 1997). Jury decisions also occasionally foreshadow actual legal reforms (e.g., *R. v. Morgentaler*, 1988). There are at least two dimensions that comprise negative insanity related attitudes: one concerning myths about the defence, and one relaying notions of strict liability (i.e., "If you do the crime, you do the time", Skeem et al., 2004). Hence some laypersons are simply recalcitrant on the matter of mental disorder and criminal responsibility.

Granted, it is not a truism to label the principle of strict liability an error,

especially because we cannot easily establish ground truth in every case. Scholars have, after all, long debated the role that 'commonsense' justice should have in informing legal policy (Finkel, 2000; Haney, 1997), and some question the ethics of the insanity defence. Notably, such debates are beyond the scope of this thesis. Nonetheless, from an empirical standpoint, corrections centers are arguably not the best option. Persons with mental disorders are not only over-represented in the penal system, but evidence abounds that corrections centers are ill-equipped to provide appropriate mental health services (Adams & Ferrandino, 2008; Olley, Nicholls, & Brink, 2009; Slinger & Roesch, 2010). Although many professionals have sought means to divert those with mental disorders away from traditional punishment (Slinger & Roesch, 2010), bias against the insanity defence among the public may add to the problem. While I do not dispute the import of commonsense justice, I leave prescriptions about the law to other disciplines. Research clearly shows, though, that misconceptions about the insanity defence contribute to lay opinions to some extent (Bloechl et al., 2007; Maeder et al., 2015; Skeem et al., 2004).

It further bears mentioning that the Canadian government has actually called for *stricter* treatment of NCRMD defendants in the wake of high profile cases, by way of the 2014 Not Criminally Responsible Reform Act. In the words of former Canadian Prime Minster Stephen Harper, the NCR Reform Act would: "give the courts the powers they need to keep those deemed too dangerous to be released where they should be — in custody" (Cohen, 2013). Yet, in light of the data on negative attitudes toward NCRMD, and a handful of benchmark cases (*R. v. Swain*, 1991; *Winko v. British Columbia*, 1999) that have affirmed the need for less restrictive dispositions, the Act risks compounding

the problem by discouraging insanity pleas.

Moreover, every person has the constitutional guarantee to a fair trial by an impartial tribunal (Canadian Charter of Rights and Freedoms, 1982, Section 11d). A pressing issue is therefore whether the legal system lacks adequate safeguards to combat juror partiality. If jurors are unwilling to apply the current laws, whether by reason of misinformation or moral objection, then a second course of action is to persuade jurors to use NCRMD in appropriate cases. Researchers have proposed several mechanisms by which these biases manifest, including lack of education about the insanity defence. Thus, in order to proprely process case facts, we should ensure that jurors are sufficiently educated on NCRMD. However, recent research indicates that education does not always work. For instance, Maeder et al. (2015) found that while providing mock jurors with focused legal education produced more favourable attitudes toward NCRMD, there was no behavioural difference; participants were unwilling to apply NCRMD when a case met the prongs of insanity. Even if we relieve laypersons of misconceptions about the defence, resilient beliefs that all persons must be punished regardless of mental disorder might require a different intervention.

Accordingly, Maeder et al. (2015) argued that correcting misinformation on the insanity defence is not enough – it could also be a matter of punishment orientation. Although researchers have remarked on the kinship between insanity defence attitudes and different punishment orientations (Breheney, Groscup, & Galietta, 2007; Skeem et al., 2004) existing proposed remedial measures do not target all punitive motives. For example, assuring jurors that incapacitation and deterrence are guaranteed with an insanity finding quells utility focused punishment goals, but the negative bias might also

be seated in desires for retribution. The central role of punishment orientation is therefore largely uncharted territory in the realm of insanity research. Studies from the moral judgment arena surrounding the distinction between moral intuitions and reasoning may be informative. As Dwyer (2009) summarized, much of this work can be classified as either 'sentimentalist' or 'rationalist', with some (e.g., Greene, 2009) advocating for the existence of both an automatic, emotional system, and a rational, cognitive one. It is well established that insanity myths play a key role in aversion toward the defence, but it is possible that certain moral intuitions interfere after correcting misconceptions.

Even so, from an applied standpoint, there are still limitations to the vehicle through which to educate jurors. Judge's instructions are one potential educational avenue built into the legal process. While there are model judge's instructions available (National Judicial Institute, 2014) – parts of which seem to leverage the power of education by assuaging fears about the defence – some theoretically effective passages are not always used. For instance, information reminding jurors that NCRMD offenders do not simply return to the public is included, but judges are not obligated to use them. Wheatman and Shaffer (2001) reported that including dispositional instructions has a significant effect on post-deliberation verdicts, although this does not seem to be true for pre-discussion judgments (Wheatman & Shaffer, 2001; Whittemore & Ogloff, 1995). Indeed, the jury deliberation process may itself be seen as a safeguard, and has been known to attenuate bias (Wheatman & Shaffer, 2001).

Barring success in relieving jurors of any cognitive errors, as well as utilitarian or retributive concerns, a third possibility pertains to juror selection. If a subset of Canadian citizens do not endorse the insanity defence, then perhaps they should not serve in

NCRMD cases. Yet, unlike other trial types – for instance the requirement of 'death qualification' (lacking extreme views on the death penalty) to serve on a U.S. capital jury – there is no standardized procedure for gauging jurors' motivation or ability to follow the law in NCRMD trials. However, the courts do reserve a Challenge for Cause procedure in the event that there is reasonable potential for partiality in a juror pool. In some cases, the Judge will permit lawyers to ask a handful of questions during selections in hopes to gauge juror biases. Thus, it is worth examining jurors' ability to self disclose objections against the insanity defence.

Considering the many issues that NCRMD defendants potentially face in orchestra with policy level changes, it is clear that we must probe whether negativity toward legal insanity reflects a lack of information on the part of the public or moral intolerance of the defence. At this juncture, it is also necessary to glimpse inside the jury deliberation process to fully diagnose the problem. In so doing, we can examine not just opinions about NCRMD, but also how they fare in response to persuasion. Whereas researchers have made gains in understanding individual juror decisions in insanity cases, only one (U.S.) study of which I am aware (Wheatman & Shaffer, 2001) has examined group deliberation in insanity cases.

The aim of this dissertation was twofold: one remedial and one descriptive. First, I investigated whether an insanity specific juror selection procedure could be useful in identifying bias. I also examined whether endorsement of retributive ('an eye for an eye') and utilitarian ('for the greater good') principles to support punishing versus avoiding punishment affected juror decisions. Second, I explored themes that emerged in jury deliberation. On a broader level, I hoped to gauge commonsense notions of justice

surrounding the insanity defence in Canada.

To begin, I provide a brief history of the insanity defence in Canada. Next, I lay a foundation by describing the psychology of attitude formation and moral judgments, paying particular mind to the distinction between moral reasoning and moral intuitions. I then provide some general information on the jury system, individual juror decision-making models, and the influence of groups on decision-making. I also discuss potential remedial measures including Challenge for Cause and judge's instructions. Lastly, I describe three studies: an examination of individual juror attitudes; a quantitative, theory-driven content analysis of mock jury deliberation sessions in a fabricated murder case; and finally, a qualitative description of the relevant content.

## Insanity

**Mental disorders**. Insanity is a legal rather than psychiatric term. To determine if a defendant is legally insane, it is the jury's duty to consider:

> (1) if it is more likely than not that the defendant was suffering from a mental disorder at the time of the act, and (2) if it is more likely than not that the defendant's mental disorder made him/her incapable at the time either of appreciating the nature and quality of the act or knowing the act was wrong (National Judicial Institute, 2014).

The law defines a mental disorder as follows for the jury: "any illness, disorder, or abnormal condition that impairs the human mind and its functioning" (National Judicial Institute, 2014). If the jury unanimously answers 'yes' to both of these questions, then the defendant is considered to lack the required mens rea for criminal responsibility. A defendant is always presumed to be innocent until the Crown has proven beyond a

reasonable doubt that he or she is guilty. Mental disorder is an exception; the party raising the NCRMD claim must prove that it is 'more likely than not' that the defendant had a mental disorder at the time of the crime. The standard of proof is lower than beyond a reasonable doubt, which is articulated in the model instructions as well.

The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (American Psychiatric Association, 2013) – which now relies on a single axis system, in contrast to the mutliaxial system that separated psychological disorders, personality disorders, and medical conditions – lays out the necessary criteria. Psychiatrically speaking, we can generally define a mental disorder as:

> A syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning. (DSM 5, 2013, p. 20)

It is important to acknowledge the ongoing debate about appropriate terminology when discussing this topic. Some argue that terms such as "disorder" and "illness" pathologize what would be better termed "mental health conditions." Others prefer that these conditions are recognized in the same way as other physical illnesses. Similarly, appropriate legal language has featured heavily in reforms to the defence in both Canada and the United States. In Canada, many objected to the term "not guilty by reason of insanity," which resulted in what we now call the NCRMD defence. Throughout this dissertation, I mirror both the DSM and the law in using the term "disorder". Although the word "insanity" no longer features in Canadian law, I retain it as an umbrella term

that refers to mental disorder in a legal context, in line with the NCRMD literature.

Data from the National Trajectory Project on Individuals Found Not Criminally Responsible on Account of Mental Disorder in Canada (NTP, Crocker et al., 2015) showed that 70.9% of primary diagnoses at the index NCRMD verdict were psychotic spectrum disorders (e.g., schizophrenia) while 23.2% were mood disorders. Consequently, in this dissertation I focus on juror decision-making in an NCRMD case involving schizophrenia, and so a brief description is needed. Namely, to meet the diagnostic criteria of schizophrenia under DSM-5, a person must experience, for a minimum of one month, at least two of the following: delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behaviour, or negative symptoms such as lack of emotion. These symptoms would also produce marked distress to social/occupational functioning. It is noteworthy that persons with schizophrenia seem to be subject to a high degree of stigma, wherein laypeople are skeptical of the potential for treatability and associate the illness with dangerousness (Angermeyer & Dietrich, 2006; Day, Edren, & Eshleman, 2007). Therefore, this group might be especially vulnerable to insanity defence bias. Given a handful of violent cases heavily reported on in the media (e.g., the 2008 Greyhound Bus incident), it is also possible that schizophrenia would serve as an NCRMD prototype in jurors' minds, and so it is a useful point of departure.

**History of the insanity defence**. As intense media coverage foreshadows, the insanity defence has a contentious history in Canada, the United Kingdom, and the United States. Following the 1843 case of *R v. M'Naghten*, the British House of Lords found the defendant Not Guilty by Reason of Insanity (NGRI), thus establishing "a defence on the ground of Insanity" owing to a "disease of the mind". Canada followed

suit and implemented the insanity defence in 1892. While the courts still rely on the test arising from M'Naghton – outlined in the previous "Mental Disorders" section – there have been a handful of benchmark cases that influenced the treatment of NCRMD accused with respect to the potential for indeterminate detention (Penney, Morgan, & Simpson, 2013). These modifications were put forth in Bill C-30, which came into effect in 1992. Owing chiefly to the case of *R. v. Swain* (1991), amendments featured a change to the language "not guilty by reason of insanity" in favour of "not criminally responsible on account of mental disorder." In particular it was noted that indeterminate detention was a violation of the Charter of Rights and Freedoms.

The decision in the case of *Winko v. British Columbia* (1999) was further influential in dictating which dispositions are per se appropriate. Following that decision, the courts maintained that unless the accused is shown to be a significant threat to society (of which there must be certainty), he or she is entitled to an absolute discharge (Penney et al., 2013). Currently, when a defendant is found NCRMD, an independent Review Board is responsible for deciding upon the appropriate disposition: absolute discharge, conditional discharge, or detention in hospital (Charette et al., 2015). In light of an increase in the number of accused found NCRMD, researchers speculate that these revisions have prompted greater defendant willingness to use NCRMD owing to the improved protection against Charter violation (Penney et al., 2013; Roesch, Ogloff, Hart, Dempster, Zapf, & Whittemore, 1997). More recently, in four major studies, the National Trajectory Project (Crocker et al., 2015) undertook to find an accurate portrait of the insanity defence in Canada. To do so, they examined archival data of 1800 men and women found NCRMD between 2000 and 2005 in three provinces (the full population

for British Columbia and Ontario, and a random sample by region from Quebec, see Charette et al., 2015). One major finding was confirmation that contrary to public perceptions of the NCRMD defence, violent offences make up only a small minority of all NCRMD cases (Charette et al., 2015).

The United States features a similar treatment of mental disorder and criminal responsibility – showing similar patterns of usage (Vitacco, Malesky, Erickson, Leslie, Croysdale, & Bloechl, 2009) – with some key differences. First, the availability of the insanity defence as well as the standards for NGRI vary across states. Most follow the M'Naghten Rule, while others use the American Law Institute's standard, which requires the defendant to have been unable to appreciate the criminality of the act or conform to the law (Sloat & Frierson, 2005). Further, some states allow for a verdict of Guilty but Mentally Ill (GBMI), which recognizes that even if a case did not meet the prongs of NGRI, mental health care is sometimes required. Researchers in Canada and the U.S. alike have focused intently not only on actual plea rates, but also on the public's seemingly incongruent attitudes surrounding the insanity defence.

**Attitudes toward the insanity defence**. Skeem and colleagues (2004) produced a reliable and validated measure of lay attitudes toward the insanity defence (the Insanity Defence Attitudes-Revised scale, IDA-R), which has shown predictive utility for mock jurors' case judgments. The IDA-R measures insanity defence attitudes across two dimensions: strict liability and perceived injustice and danger. Strict liability refers to the notion that mental disorder is irrelevant to the crime. Perceived injustice and danger refers to the belief that the insanity defence is misused and is a threat to public safety. Maeder et al. (2015) confirmed this factor structure in a Canadian sample and found that

insanity defence attitudes were largely negative and significantly related to verdict decisions in fabricated NCRMD cases. Bloechl et al. (2007) similarly found that holding misconceptions (e.g., overestimating the frequency of insanity pleas) related to negative attitudes toward the defence.

Some members of the public also erroneously believe that defendants who are found NCRMD are released into the community without provisions. Those who estimate that insanity defendants immediately go free are less likely to support the insanity defence and are more likely to vote guilty (Skeem et al., 2004). Notably, some work has probed potential gender differences in endorsement of the insanity defence. Specifically, Breheney et al. (2007) concluded that 'gender matters' in insanity cases after examining the influence of juror and defendant gender for six different mental disorders. Their data supported previous work illustrating women's greater acceptance of psychiatric testimony (e.g., Faulstich, 1984). As predicted, Breheney et al. (2007) showed that women attributed less responsibility to a defendant with a mental disorder compared to men and were also more likely to indicate that this disorder caused the target's behaviour. In short, one way to encourage appropriate use of an NCRMD finding is to correct misconceptions.

Some researchers have proposed the use of focused education to combat bias against defendants who are reasonably using the insanity defence. Indeed, Hans (1986) showed that those with higher levels of education in general are more likely to support the insanity defence. Similarly, Maeder and Laub (2012) found that psycho-legal education (i.e., an undergraduate Psychology and Law class that featured lectures on the insanity defence) improved student attitudes toward the defence. In two studies, Maeder

et al. (2015) attempted to educate Canadian mock jurors about the NCRMD defence in hopes that it would improve relevant attitudes. In the first study, education produced the predicted difference in NCRMD attitudes, but it did not affect verdict decisions. In contrast, the second study revealed no such effect of focused education on attitudes, or verdict decisions. Of note, they always measured attitudes after presentation of the trial stimuli and used different case facts across the two studies. Hence, education is not always an effective remedy to negative insanity defence attitudes. Maeder et al. (2015) also tested for the influence of gender on education effectiveness but observed no statistically significant differences. In considering the reasons why focused education did not successfully alter participants' decisions in NCRMD cases, we must turn to theories of attitude formation as well as theories of moral judgments.

**Attitudes**

Allport (1954) charaterized prejudice as an inevitability of human interaction. Confronted with a wealth of information in everyday life, people tend to rely on mental shortcuts by categorizing objects according to similarity (Fiske, 1998; Kahneman & Tversky, 1979; Van Bavel & Cunningham, 2010), which in turn direct appraisal of and behaviour toward those objects. In general, attitudes can be regarded as an orientation to evaluate other people, things, events, and ideas as good or bad (Banaji & Heiphetz, 2010). As Allport (1954) espoused, in order to combat the "antipathy based on faulty and inflexible generalization" that comprises prejudice, we must target both cognition and affect (Jackson, Hodge, Gerard, Ingram, Ervin, & Sheppard, 1996). Accordingly, to understand attitudes, researchers have advocated consideration of the 'ABC's': affect, behaviour, and cognition (Abelson, Kinder, Peters, & Fiske, 1982; Millar & Tesser, 1986;

Rosenberg & Hovland, 1960). For instance, we may regard stereotypes (perceived shared group attributes) as cognitive beliefs about an attitude object, prejudice as affective (i.e., providing valence for that belief), and discrimination as a behavioural manifestation of these tendencies.

A rich history of social cognitive research has demonstrated the many errors to which people can succumb. When evaluating an appropriate insanity claim, the jury's task may be interrupted by faulty cognitions such as stereotypes or misinformation. Repeated media representations of persons with mental disorders as dangerous and criminal might maintain stereotypes. As we have seen, jurors seem to have some preconceived associations between schizophrenia and dangerousness (Day, Edren, & Eshleman, 2007). Such tendencies may be attributable to a mental shortcut referred to as the 'representative heuristic', in which people are judged according to their similarities with existing mental prototypes of a subject (Kahneman & Tversky, 1979). People might, for instance, have an existing prototype of an insanity defendant as a malingerer, given widespread misinformation about the defence. Jackson et al. (1996) noted the further distinction that affectively driven attitudes tend to relate to direct contact with an attitude object (e.g., knowing a person with a mental illness) whereas stereotypes tend to follow from indirect contact (e.g., the media). Of course, people could for any number of reasons avoid individuals with mental disorders, whether an antecedent or consequence of stereotypes and prejudice.

There is some disagreement as to whether behavioural intentions toward an object constitute attitudes, given the documented lack of attitude-behaviour consistency (Millar & Tesser, 1986). Eagly and Chaiken (2007) urge researchers to distinguish between

evaluative responses (e.g., verdict decisions in the current studies) and attitudes themselves, viewing the former as an expression of attitudes. In the context of this dissertation, beliefs about common practices (e.g., the NCRMD plea rate) can be characterized as cognitive, while concerns about the dangerousness of defendants can be viewed as affective. It is instructive to consider whether strict liability is cognitive or affective. The belief that a crime per se requires punishment could be regarded as a rational calculation or conversely as an affective evaluation of what is right. Research from the moral reasoning literature as well as behavioural neuroscience sheds light on such questions.

**Rationalist and Intuitionist Theories**

The increasing popularity of contrasting emotional and cognitive processes in decision-making follows an unresolved, centuries-old debate as to the true nature of moral judgments (e.g., Hume; 1738/2000; Kant, 1785/2006). In the parlance of cross-disciplinary literature, there are two conflicting camps: rationalist theories and sentimentalist or intuitionist theories (Dwyer, 2009). Rationalist theories conceptualize moral reasoning as a methodical process in which people reason through available evidence in coming to a conclusion. Conversely, intuitionist theories hold that moral leanings are a precursor to reasoning, followed by retrospective justification.

**Moral reasoning.** Rationalist theories view moral judgments as the result of explicit cognitive reasoning. Thus, from one perspective, it is possible that people appraise available information about the cause or intentionality of an event and then match their decision accordingly. As Weiner (1997) observed, mental disorder should in theory operate as a mitigating factor when engaging social cognitions about an event,

eliciting sympathy rather than anger. Following a taxonomy of causality, Weiner (2006) deconstructed the way that people try to judge a positive or negative occurance (i.e., the type of attributions they employ). He described the three dimensions across which people think about the origins of the incident: control, locus of causality, and stability. If an observer determines that the cause of an event was internal (as opposed to environmental), controllable, and stable, the theory suggests that responsibility and hence blame and anger will result.

Relatedly, researchers have also observed that people react differently to issues surrounding how a condition was brought about and steps that could be taken to remedy the condition (Brickman, Rabinowitz, Karuza, Coates, Cohn, & Kidder, 1982; Weiner et al., 1988). For instance, Weiner et al. (1988) showed evidence that physically-based illnesses such as cancer tend to be seen as uncontrollable, whereas people attribute control to those with mental-behavioral issues such as substance abuse. In essence, it seems that mental-behavioral disorders are typically labeled 'moral weaknesses' (Weiner et al., 1988). However, it is worth noting that people vary in their perceptions of schizophrenia; some might view it as controllable (e.g., through medication), and others might recognize the ways in which the disorder itself interferes with medication compliance (Martin, Pescosolido, & Tuch, 2000; Yamamoto et al., 2017).

It is also possible that people make causal attributions according to the moral status of an event. The Knobe effect **(**Knobe, 2003**)** illustrates that people sometimes judge the intentionality of an act on the basis of its valence alone. For example, participants provided with a vignette describing a CEO who decides to implement a profitable program despite anticipated harm to the environment tended to ascribe

intentionality, whereas those who read that it would help the environment did not. More broadly, researchers have described a fundamental attribution error (Ross, 1977), by which people as observers tend to rely on dispositional explanations when understanding negative events, and ignore contextual information. Pettigrew (1979) argued that this error is amplified when the observer sees the target as 'other'. In this context we might expect a person to distance him or herself from one who has committed a moral violation. If, as the Knobe effect (Knobe, 2003) and attribution errors (Ross, 1977; Pettigrew, 1979) imply, the moral status of an event can dictate one's explanation for that event, then we might suspect juror judgements to be more intuitively based. That is to say, mock jurors might have a proclivity to attribute blame by virtue of the negative event itself. However, it is necessary to probe individual differences to better predict for whom this might be the case. Moreover, it appears that there are many potential sources of information and error for jurors' appraisal of case facts, having both affective and cognitive origins. Surely, reliance on only one intervention strategy will only be partially effective.

**Moral foundations theory.** A frontrunner in the intuitionist perspective, Haidt (2001) has maintained that evaluative feelings about another's actions defy easy articulation, and emerge automatically without consciously weighing rational premises. One salient example is the common aversion to a romantic relationship between siblings (Haidt, 2001). People seem to have a sense that it is wrong even if they cannot justify that feeling, a phenomenon known as 'moral dumbfounding'. A key tenet of this social intuitionist account is the assertion that moral 'reasoning' is retroactive (Saltzstein & Kasachkoff, 2004). Haidt and Graham (2007) proposed that there are five moral intuitions (also called foundations): fairness/reciprocity, harm/care, authority/respect,

sanctity/purity, and ingroup/loyalty. They posited that these culturally invariant, inborn ideas operate beneath conscious awareness and are derived from evolutionarily beneficial practices such as reciprocal altrutism (helping another with the expectation of later help).

Each foundation may be rejected or accepted as a basis for moral virtues (i.e., qualities that make a person 'good'); different cultures sometimes emphasize different virtues. The fairness dimension concerns notions of equality and equal protection. For instance, people tend to be concerned with the justness of the procedures used to make decisions, sometimes more so than with the actual outcome (Tyler, 1984). The harm dimension concerns preference for actions that promote safety rather than suffering. For instance, some may find it virtuous to prevent the suffering of the greatest number of people, while others cannot tolerate harm to a single individual (Foot, 1967). The authority dimension concerns deference to hierarchy. For instance, some cultures may emphasize subordination whereas others value challenges to authority (Haidt & Graham, 2007). The sanctity dimension extends more primitive notions of uncleanliness and danger to the moral realm. For example, religious virtues may dictate appropriate bodily activities such as sex (Haidt & Graham, 2007). Finally, the ingroup dimension concerns the natural tendency to socially categorize others and preference loyalty to those perceived as similar.

**Dual-process theories**. Evidence from neuroscience bolsters the claim that judgments may have affective or cognitive origins. According to Greene's (2009) dual process theory of moral reasoning, there are two routes by which people make judgments. When making so called 'deontological' decisions pertaining to a person's fundamental rights (e.g., not to be harmed), emotion structures of the brain are active. Another 'cold'

system, on the other hand, related to cognitive processing is active when one is making rational judgments about the greater good. Further, Bartels and Pizarro (2011) found that absent normal emotional processing, individuals scoring high on psychopathy and Machiavellianism are more apt to endorse tasks that would require a moral violation (e.g., the death of another).

The dual process account (or single process in some interpretations of those data, Bluhm, 2014) acknowledges the role of both emotions and cognitions. Earlier writings on dual process (Greene, 2009) relied on the idea that information can be filtered through a more ancient, emotionally-driven system that allows for quick appraisal of stimuli, as well as a higher-order one that can override it (Goleman, 1995). By way of illustration, Greene (2009) likens the process to a 'dual-mode camera', which allows for both automatic and manual operation. As Bluhm (2014) pointed out, some cognitions and emotions work in orchestra, while in other circumstances they compete. Feigenson and Park (2006) defined emotions as inclusive of "feelings, cognitions, and actions, or inclinations to act" that "help that person choose among and coordinate competing goals and values" (p.144). Feigenson and Park (2006) further emphasized the widely accepted role of cognitions in theories of emotion, which aid in appraising the significance of the change to one's environment that an emotion signals. Greene's (2009) data show that judgments for scenarios involving more direct harm to a person (e.g., pushing him off of a footbridge) tend to occur through systems that are also engaged during emotional reactions, while indirect harms rely more on 'cold' systems. Bluhm (2014) argued that cognitive systems are simply a means to compare different outcomes; emotions provide information on the valance of various outcomes to be weighted. While the debate

between intuition and rationality has raged on, until more recently the role of emotions has seen comparatively little attention in the psycholegal realm (Wiener, Bornstein, & Voss, 2006). In brief, researchers describe emotion as having three general effects on decision-making: by dictating information processing strategy, by biasing judgments in congruence with one's mood, and by offering informational cues to responsibility attributions (Feigenson, 2006).

**Punishment Orientation**

Whether one subscribes to a rationalist or intuitionist account, it is clear that guilt and punishment decisions are inextricably linked. In an insanity case, the actus reus is often established. By virtue of the fact that jurors are instead assessing the defendant's control over the act, labeling this as a 'guilt' decision is somewhat of a misnomer. Rather, jurors will dictate whether the defendant is to be punished or treated, which is at least in part a question of punishment orientation. The system intends for jurors to only rely on evidence of whether the defendant had a mental disorder at all (and whether it precluded mens rea), but this is not to say that jurors will avoid retrospective justifications. Studies seem to suggest that people do not always reason through the evidence (Haidt, 2001).

Further, recent work illustrates the possibility that when thinking about punishment, some people are generally punishment motivated, while others focus on the risks associated with punitive acts. To provide more background, punishment can be defined as a "a negative sanction intentionally applied to someone perceived to have violated a law, rule, norm, or expectation" (Vidmar & Miller, 1980, p.568). People tend to rely on two main types of arguments when punishing others: retributivism and utilitarianism. Retributivism, which follows from Kant's (1785/2006) Deontology, holds

that one must only act in such a way that would be permissible for all others to do the same in those circumstances. This principle is known as the 'categorical imperative.' The mainstay of this theory is that it pertains to "the right" insofar as consequences should not dictate morality (Kant, 1785/2006). Punishment must be proportionate to the wrongdoing (Schedler, 1980). Unlike retributivism, the cornerstone of utilitarianism – commonly associated with Bentham (1789/2006)**,** and later championed by Mill (1859/2008) – is consequentialism. That is, an act must maximize the aggregate good for those affected by it. Therefore, incapacitation, rehabilitation, and specific/general deterrence are utilitarian punishment practices. While a retributivist is not per se opposed to these goals, they are likely to preference balancing the scales of justice. Laypersons do not always experience the distinction between these viewpoints. In fact, Canada's sentencing principles refer to both concepts in the Criminal Code. For instance, while one objective is to "deter the offender and others from committing the same crime" (718), another requires that a sentence be "proportionate" to the seriousness of the crime and responsibility of the offender (718.1). Yet in theory deterrence often requires disproportionate punishment.

In previous work (Yamamoto & Maeder, 2019) we sought to rectify the apparent difficulties in operationally defining punishment orientation by creating four scales that measured which retributive and utilitarian principles work in tandem versus in tension. This work was based on moral psychology research showing that more logically calculated decisions tend to require suppression of an automatic aversion to doing harm (Greene et al., 2001; Valdesolo & DeSteno, 2006). However, research indicates that some people do not seem to experience this aversion (Bartels & Pizarro, 2011). In short, perhaps people differ in the extent to which they are encumbered by an avoidance of

doing harm. Hence, we theorized that people would differ in terms of seeing punishment as itself rewarding (Yamamoto & Maeder, 2019).

The key finding of our studies was in favour of a non-dichotomous treatment of the two camps that allows for overlap in the psychological functions that these orientations fulfill (Yamamoto & Maeder, 2019). Retributivism and utilitarianism were each branched into 'permissive' and 'prohibitive' constructs. The label 'prohibitive' connotes that the orientation is punishment-averse, whereas the label 'permissive' connotes that the orientation is punishment-prone. A prohibitive retributivism construct concerned a desire to ensure that innocent persons are never punished and captured a person's focus on the risks associated with punishment. A prohibitive utilitarianism counterpart captured the belief that in order to be justified, a punishment should yield some benefit and should be goal-oriented. A permissive retributivism dimension was characterized by blame of the offender. There was also a permissive utilitarianism dimension, which concerned endorsement of disproportionate punishment to ensure deterrence and public safety.

In the context of the current studies, if we view the purpose of legal sanctions as denouncing immoral behaviour and protecting the public, then sending (non-malingering) NCRMD offenders to the mental health care system should be considered an appropriate outcome. It stands to reason that those who are less punishment-prone may be more open to educational interventions that target insanity myths. Retributivism is by definition retrospective inasmuch as the decision is only needed by virtue of a violation (Tebbit, 2005). A retributive punishment orientation is already begging the question; because there is no requirement of future utility, when one uses retributive punishment it is

because he or she wants to punish (Tebbit, 2002). By way of example, death qualified jurors tend to also be prosecution/guilt prone (Butler & Moran, 2002; Fitzgerald & Ellsworth, 1984; Young, 2004). In other words, the individual difference (independent of the case) of willingness to take a defendant's life relates to endorsement of *guilt*. When a case is severe, it seems that retributivism may take the wheel. This notion lends to an intuitionist view on moral judgments, bolstered by previous findings on the relationship between deontology and automatic, emotional moral processing (Greene, 2009).

Consider that jurors are first presented with a transgression of certain severity, one that the Crown – the authority and gatekeeper of criminal charges – has officially leveled against the defendant. The mere fact of an indictable offence has potentially triggered a punitive need (that must be sated) to which jurors may match the incoming information to the least dissonant story. The legal system generally discourages jurors from considering the consequence of a verdict (Wheatman & Shaffer, 2001). Many have argued that attention to sentencing distracts from their fact-finding duties (see Wheatman & Shaffer, 2001). However, the fact that jurors are asked to ignore the possible punishment does not guarantee that they are able to. The concept of strict liability deals directly with the outcome of insanity trials. Endorsing such a viewpoint implies that prison is preferred over institutionalization. If we take retributivism to mean deserving of punishment only when accompanied by intentionality, then prison is incongruent. However, if we take retributivism to mean deserving of punishment to balance the scales of justice (i.e., by virtue of a harm), then prison may be desired.

NCRMD decisions are a natural moral conflict. Jurors are presented with an incident of harm, and yet told that it is not necessarily punishable. Hence there is likely a

third unspoken question inherent to jurors' assessments: whether NCRMD satisfies the goals of punishment. In Studies 1 and 2, I tested whether participant scores across the permissive and prohibitive dimensions predicted verdict decisions and deliberation content, as well as endorsement of the insanity defence. While there is ample reason to predict that punishment orientation relates to individual insanity judgments, observing how jurors of varying orientations attempt to persuade each other may be especially diagnostic. After all, the very point of the jury is to gather a sufficient diversity of opinions so as to get closest to the truth. The sections to follow contain more information about the jury system. In light of the pivotal role of group discussion in the legal system, I also describe social psychological findings that bear upon juror and jury decision-making.

**Juries**

In order to reap the benefits of the adage 'twelve heads are better than one', the Supreme Court of Canada holds that juries must be representative and impartial. Ontario jurors are drawn from the Municipal Property Assessment Corporation (MPAC; Ministry of the Attorney General, 2015). Once a citizen receives a jury summons, he or she is required by law to attend jury selections except under certain conditions, such as when jury service would produce unnecessary hardship. Medical practitioners, law enforcement, Armed Forces, legal students and practitioners, and other government officials do not qualify. Jurors generally go home at the end of each day, with the exception of criminal trials in the event that deliberation has begun and they have not reached a decision. In Ontario, whether or not a juror is selected, he or she is ineligible for three years following service. Unlike in the U.S., Canadian juries in criminal trials

must have 12 members, allowing for a minimum of 10 when jurors cannot continue service. In some cases a judge may consider it necessary to retain one or two alternate jurors. The enterprise of the jury system is not unlike a scientific experiment. Reasonable attempts are made to get a random sample that is most representative of the population. Likewise, the jury room is an amalgamation of social psychological phenomena in an applied and consequential setting; therefore, it is ripe for empirical investigation.

**Jury decision-making research.** The Chicago Jury Project (Kalven & Zeisel, 1966) is often regarded as the inception of jury decision-making (JDM) research. Analyzing data from 3576 criminal trials, Kalven and Zeisel (1966) made the first systematic attempt to uncover the number of jury trials in the U.S. among various other properties (e.g., the likeliness of hung juries). Since Kalven and Zeisel's (1966) famous finding that a first ballot majority vote predicted verdicts an impressive 90% of the time, jury literature has focused heavily on individual decision-making, absent any group discussion component. However, as Salerno and Diamond (2010) reminded us, while Kalven and Zeisel's familiar metaphor (likening deliberation to a photo development process from exposed film) is still touted as gospel, this may underestimate the influence of group dynamics. A bounty of social psychological research clearly forecasts significant group phenomena. More recently, Devine (2012) proposed a multilevel model of jury decision-making that accounts for both individual and group processes.

The study of individual juror decision-making can be divided roughly in two: mathematical and explanation models. Mathematical models dominated the earlier years of JDM research, which held that one could algebraically weigh various pieces of evidence so as to produce a numerical verdict probability. As Devine (2012) summarized,

a main distinction between such models is that whereas more descriptive models seek to understand what jurors do in reality, prescriptive models rely on logical premises about how information should be influential. While precise and quantitatively testable, researchers have criticized the mathematical method for being an unrealistic portrayal of how jurors process trial information. Instead, many have argued that jurors are not passive receivers of data but actively create meaningful narratives. In particular, Pennington and Hastie (1986) outlined a three-stage "story model" for understanding the decision-making process of individual jurors. First, the juror forms the presented evidence into a story. Second, following instructions from a judge, the juror categorizes the verdict choices. Finally, the juror attempts to match the acceptable story to an appropriate verdict.

Devine (2012) has since expanded this model to account for the fact that jurors' individual stories are built upon an existing foundation of prior case knowledge, personal beliefs, and information coming from opening statements. Relying on a film director metaphor, the "Director's Cut Model" holds that jurors' exposure to case-specific factors (e.g., seriousness), defendant characteristics (e.g., having a criminal record), and the jurors' own characteristics (primarily race, gender, socioeconomic status, trust in the system, and the Need for Cognition) initially contribute to the "shooting script" upon which they will make changes as they receive information. He theorized that jurors will then consider each piece of the defence and prosecution evidence across three domains. They will think about the scope (richness of evidence), credibility, and singularity (convergence on a single story) of the prosecution's case. The defence can offer a number of alternative stories, asserting errors to: the defendant (i.e., was not present at the crime),

actions (i.e., was present but did not commit the crime), the mindset (e.g., unintentional), and consequences (i.e., no actual harm resulted; Devine, 2012). Jurors receive this information to the aim of four possible cognitive states: believers (favouring the Prosecution's story), doubters (favouring an alternative or Defence story), mullers (vascillating between stories), or puzzlers (unable to form a story). The next step to understanding the decision-making process describes the effects of group deliberation, referred to as the Story Sampling Model. Before detailing this model, background on seminal group decision-making studies is helpful.

**Group influence**. In a hallmark study on the power of conformity, Asch (1951) investigated the extent to which individuals would bring their judgments in line with others. Participants were first shown a reference line and then asked to indicate which of three other lines (one shorter, one the same, one longer) was identical to the reference line. The task was created such that the rate of correct responding should be virtually perfect. Participants said their answers aloud after three confederates answered first. Asch (1951) found that in 37% of the trials in which the confederates indicated the same wrong answer, the participant conformed by also indicating the wrong answer. However, when one confederate gave the right answer (i.e., the participant had an ally), conformity dropped to only 5%. Yielding was also much lower (by two thirds) when the participant was permitted to provide a written rather than public response.

As Deutsch and Gerard (1955) pointed out, participants in Asch's (1951) study were not necessarily made to feel like functioning members of the group, but rather simply made judgments in the physical presence of others. Thus, Deutsch and Gerrard (1955) distinguished between two forms of social influence: normative and

informational. Normative social influence features a desire to align with positive expectations of the group, whereas informational social influence features acceptance of information from others as 'evidence of reality" (Deutsch & Gerrard, 1955, p.629). In the Asch (1951) conformity study, some participants believed that they must simply be making an error, whereas some knowingly provided the wrong answer so as not to go against the others. To investigate this distinction, Deutsch and Gerard (1955) replicated Asch's (1951) study with a few key changes. In one variation, rather than being face-to-face, participants were separated by a partition, indicated their answers via a button, all participants were naïve, and confederates were 'electrical' (i.e., a manipulated lit indicator panel displayed the same pattern as in the Asch experiment). The 'group situation' variation followed the same procedure, except that participants were instructed that they were part of a group of similar others and the best five groups would be rewarded. The addition of this group goal significantly increased conformity, demonstrating the role of normative influence.

The results of conformity studies herald different outcomes dependent on several features of the jury. Most notably, these findings bear upon effects of jury size. The U.S. Supreme Court has held that juries as small as six are functionally equivalent to twelve person juries (*Colgrove v. Battin*, 1973; Devine, 2012; *Williams v. Florida*, 1970) – although Canada has not followed suit. Yet, Asch's (1951) comformity study demonstrated that the presence of an ally can dramtically reduce conformity, and therefore six person juries are less likely to produce the necessary distribution (Hastie, Penrod, & Pennington, 1984). Among the studies that have tested for differences between six and 12 person juries, no reliable effects have emerged (Hastie et al., 1984; Devine,

2012). Still, there is evidence indicating that larger juries are more likely to hang (Kerr & MacCoun, 1985). Saks and Marti's (1997) meta-analysis also showed that there is a greater likelihood of having a minority member on larger juries, and other research demonstrates that racial heterogeneity appears to produce greater quality deliberations (Sommers, 2006).

There are several influential factors aside from jury size. Social decision schemes (SDS) provide useful rules or models that govern how individual decisions converge to produce group outcomes. Most notably, Davis (1973) sought to test several hypotheses in which individual preferences could be mapped out in a matrix to yield a probable outcome. For instance, Davis (1973) considered Lorge and Solomon's (1955) observation – that if one or more group members propose a a correct response then the group will accept it – to be a 'truth wins' SDS. Simialrly, 'majority rules' is an SDS that relies on the notion that minority faction members yield to the group (Davis, 1973). Social Transition Schemes (STS) add a description of changes to verdict distributions throughout deliberation, for instance, as a function of length of time permitted and polling method (Devine, 2012).

Polling style can indeed have a significant impact on jury decision-making. Juries tend to feature some form of informal vote during deliberation (Devine, 2012), whether by show of hands one by one, all at once, or by anonymous piece of paper. As Devine (2012) described, whereas some juries tend to center discussion around choosing a verdict (with strong advocates for each side), others focus discussion around case facts and tend to take later polls. The latter kind (referred to as 'evidence-driven' juries) form a narrative out of the case facts rather than initial feelings about the verdict. Research

seems to converge on the finding that, in the interests of justice, it is preferable for a jury to be evidence- rather than verdict-driven (Devine, 2012). Based on post trial U.S. juror self reports – which are illegal in Canada – of the timing of the first poll, the highest estimate of verdict-driven juries was 50% (Ellsworth, 1989), with some studies yielding rates between 6% and 31% on the lower end (Devine, 2012; Devine et al., 2004; Sandys & Dillehay, 1995). Hastie suggested that roughly 38% are likely a mix of evidence- and verdict-driven characteristics. Salerno and Diamond (2010) reminded that even where a juror suggests an initial vote, it is likely to be interrupted by discussion (Diamond & Casper, 1992). When it comes to emergent verdict factions following a vote, there are also some data on the effects of size. In general, larger factions have larger sway (Devine, 2012). MacCoun and Kerr's (1988) meta-analysis showed that majority faction size (specifically double the minority faction) predicted final verdicts. Salerno and Diamond (2010) further cautioned that, without a record of pre-deliberation verdicts, if a straw poll is not immediate then it is not possible to tease apart the effects of deliberation and initial position.

Again bearing in mind the Asch (1951) conformity study, it is also clear that the order of a vote can be significant. Davis, Stasson, Ono, and Zimmerman (1988) presented mock jurors with a video trial of a fabricated assault case and took individual verdict preferences. They then randomly assigned three guilty verdict participants and three not guilty verdict participants to deliberate in groups two days later. The groups were instructed to come to a unaninmous decision, and the experimenters had groups either poll at the oustet or to wait five minutes into discussion. Votes were either simultaneous or sequential. Unbeknownst to the participants, the sequential votes were rigged to occur

with the first three members of the guilty (or not guilty) faction first. The researchers were interested in any change from initial verdict preference. Davis et al. (1988) used fourth voter changes as a proxy for sequence effects, because that participant heard three consecutive guilty or not guilty verdicts prior to voting. For early polls, participants were more likely to change from not guilty to guilty in sequential than in simultaneous polls. However this was not the case for the later poll conditions. Davis et al. (1988) noted the likelihood that social norms (and legal obligations) that necessitate greater leniency became more salient during group discussion, which accounts for the lack of yielding to a guilty verdict for the fourth voter.

Results also indicated that participants who took part in a simultaneous vote were more likely to switch to not guilty than to guilty. This finding demonstrates an important concept in jury decision-making: the tendency to shift toward leniency following group discussion (Davis et al., 1988). Given that the court sets the burden of proof intentionally high, group leniency is a desirable effect. They found no significant differences in change from guilty to not guilty when comparing the sequential and simultaneous voting groups, nor in terms of early or late polling.

Also of note, Mize et al. (2007) made the most recent attempt to estimate the actual amount of time that juries take to deliberate by surveying judges and attorneys in 11,752 trials. They reported that a basic (low legal complexity, 12 jurors with a unanimity requirement) state court civil trial had a mean deliberation time of 2.77 hours. Mize and colleagues (2007) observed a relationship between deliberation time and case type (with criminal cases that are more serious taking longer), court system (state or

federal), and case complexity, which could each add roughly one to three hours.

A final important phenomenon that can impact on deliberation is group polarization. Desribed as the 'risky shift' phenomenon by earlier researchers (e.g., Pruit, 1971), this refers to the tendency to become more extreme in initial positions following group discussion. Essentially, groups are more 'risky' than individual members. Later researchers realized that the phenomenon had broader application and was not an issue of 'risk' so much as of the influence of the group, and hence was more appropriately called 'group polarization' (Myers & Lamm, 1975).

In sum, the prevailing theory is that jurors first bring idiosyncratic interpretations of the initial case facts. Researchers have also observed some trends as a function of broader group membership (e.g., gender, race). Then, in jurors' attempts to persuade eachother, both normative and informational influences may be introduced; there are a number of social decision and social transition schemes that can affect a trial outcome, pertaining to size, polling style, and length. A central focus of the Story Sampling Model is the number and quality of stories that the jurors put forth, which range from isolated statements to complete stories (Devine, 2012). Devine (2012) asserted that the latter would have the greatest influence. In this dissertation, I relied on coding techniques that incorporated individual attitudes and group context in order to understand how these stories emerged. As laid out in the introduction, it is also necessary to investigate ways to fix any extralegal influences.

**Remedial Measures**

Canada relies on what can be called an 'adversarial' legal system. Because every person is considered innocent until proven guilty, the Crown has the burden of

demonstrating enough evidence of guilt beyond a reasonable doubt. In most cases, the Defence need not prove anything. The Defence should, however, hold the Crown to this high standard by providing the judge or jury with reasonable alternative accounts, or by pointing out gaps in their evidence. Two advocates present opposite sides of an event in the best light possible, and then a third party can best uncover the truth. Thus, the nature of our legal system is such that many of the previously discussed principles can be applied directly in practice. For instance, lawyers could simply make appeals to utilitarianism or retributivism in opening arguments. More optimistically, perhaps some jurors will naturally effect some of these remedies during deliberation, which is why it is a key next step in studying the insanity defence. However, it is likely that further interventions will be needed, and they are indeed available. Although psycho-legal researchers have admittedly limited power in persuading the courts to adopt new practices, juror selection and judicial instructions are two means to implement such remedial measures.

**Challenge for cause.** Beyond the benefits of the adversarial system, the courts have a number of pretrial procedures to identify and prevent partiality. The Canadian legal system operates on the general assumption that jurors are impartial, but under some circumstance the courts grant that this is not the case. In addition to a change of venue, an adjournment (i.e., waiting for time to pass), and peremptory challenge (i.e., initial, limited juror selection), a trial party can call for a Challenge for Cause procedure. If a trial party can demonstrate the realistic potential for partiality in a population of jurors (*R v. Sherratt*, 1991), then, at the discretion of the individual judge, lawyers may be permitted to ask some questions to gauge their biases. Two persons at a time (called ‘triers’) from

the jury pool will then judge the veracity of the responses (Canadian Criminal Code, section 638(1)(b)). Examples of successful requests for challenge for cause include suspected racial bias (*R. v. Parks*, 1993; *R. v. Williams*, 1998) and potential prejudice due to excessive pretrial publicity (*R. v. Pickton*, 2010).

Schuller, Erentzen, Vo, and Li (2015) examined Ontario selection procedures involving 1,392 prospective jurors. Tellingly, they found that while only 8.3% of prospective jurors conceded possible partiality on the basis of defendant race, more than twice as many triers deemed them unacceptable. It appears that when judging others, people consider more than direct admittance of potential bias. Of course, researchers agree that social pressures can make people unwilling or unable to self-report existing bias, especially because phenomena like racism are socially unacceptable (Gaertner & Dovidio, 2005; Paulhus, 1991).

Schuller, Kazoleas, and Kawakami (2009) have drawn attention to the problematic assumption that mock jurors are able to self-assess their own biases, understand how they might influence decision-making, and reflect on personal motivation to curb such tendencies. Indeed, Schuller et al. (2009) observed no significant differences in legal decisions between those who indicated that they were likely to be partial (in a race-related case) and those who indicated they could be impartial. This was true regardless of whether the question was open-ended (in an attempt to bolster self-reflection) or closed-ended. Notably, the process of asking mock jurors a closed-ended Challenge for Cause question did not affect bias. Perry, Murphy, and Dovidio (2015) emphasized the pivotol role of ‘concerned awareness’ in correcting for one’s biases. In the context of racism, while people have egalitarian intentions, when the rationale for a

decision is easily attributable to other sources, implicit prejudice can be influential (Gaertner & Dovidio, 2005). Nonetheless, negativity toward the insanity defence does not perfectly mirror the clear immorality of racial bias. As I alluded to in the introduction, strict liability, like death penalty support, could be considered a moral stance rather than an objective error. Hence, a selection procedure in an insanity case could potentially tap genuine beliefs about the appropriateness of NCRMD.

Another limitation to the Challenge for Cause in this context is that jurors would not necessarily be instructed on NCRMD, given that the challenge process takes place during voir dire. Sloat and Frierson (2005) investigated jurors' prior knowledge of the definition of insanity. They found that even among highly educated mock jurors, only a minority of participants could identify both the definitions of NGRI/ GBMI and the available dispositions. Tellingly, Finkel and Handel (1988) found that the presence or absence of legal instructions did not seem to result in different verdict splits. As we have seen, greater knowledge about the insanity defence would not necessarily translate to greater acceptance of it. We must first identify potential contexts or individual differences that might render accurate information about insanity more effective.

At any rate, although no one (of which I am aware) has used the Challenge for Cause in the context of NCRMD, the procedure can theoretically apply to an insanity case. Prior to the jury deliberation task in Study 1, in addition to completing the IDA-R and POQ, mock jurors were asked to indicate whether they endorse abolition of the insanity defence. This question is unlikely to be allowed as part of juror selection; it served as a comparator to the more rigid wording of the traditional question structure. In the same fashion as Challenge for Cause, participants were also asked whether they had

strong opinions about it that would preclude their ability to be impartial in a NCRMD case.

**Judge's instructions**. Once a case goes to trial, there is yet another official safeguard that can reduce partiality. Instructions regarding the NCRMD defendant's disposition have been the subject of both legal debate and empirical investigation. Because it is not the charge of the jury to consider the consequences of the law, the U.S. Surpreme Court has ruled that disposition instructions are unnecessary except in cases of serious error, for instance, if misinformation is communicated during trial (*Shannon v. United States*, 1994; Wheatman & Shaffer, 2001). However, such instructions have been allowed in at least some circumstances (e.g., *United States v. Neavill*, 1989). In Canada, the National Judicial Institute (NJI) has made a set of model instructions available to judges, which lays out standard language for instructing juries about legal rules and the jury's duties. Information about the disposition of the defendant are included and read as follows:

> Your verdict must be based solely on your consideration of all the evidence, the submissions of counsel, and the law as I have explained it to you and must not be influenced by a consideration of the consequences of a special verdict of not criminally responsible. However, you should know that our law provides that persons who are found not criminally responsible by reason of mental disorder will not be set free if they pose a significant threat to public safety. (National Judicial Institute, 2014)

Although Canadian judges are not obligated to read this particular set, as the NJI remarked, judge's instructions are key to the efficiency of the judicial process, and

therefore it would be useful to understand the import of specific passages. In the current studies this passage was therefore read to half of the juries.

Given the evidence of insanity defence myths, Wheatman and Shaffer (2001) argued that the prevailing opinion against dispositional instructions actually risks introducing error into the jury room. Wheatman and Shaffer's (2001) work followed from Whittemore and Ogloff's (1995) study of juror decisions in an NCRMD case, in which they randomly assigned mock jurors to either receive or not receive dispositional instructions. Whittemore and Ogloff (1995) did not find an effect of dispositional instructions, and further reported that few jurors could even correctly retain the information.

Wheatman and Shaffer (2001), however, took this one step further and considered the potential effect of deliberation on this outcome. They argued that because individual jurors are accountable to the group when there is a unanimity requirement, the resultant greater depth of thinking about the issues (Tetlock, 1983) yields higher likelihood of an effect of instructions. To test this hypothesis, they presented undergraduate students with a videotaped murder trial depicting a Not Guilty By Reason of Insanity case. In addition to individual verdict decisions, participants deliberated in groups of six. Replicating Whittemore and Ogloff's (1995) finding, Wheatman and Shaffer (2001) observed no differences between instructed and non-instructed pre-deliberation verdicts. However, as predicted, they observed a significant effect of dispositional instructions on post-deliberation verdicts. Specifically, whereas 60% of dispositionally instructed juries found the defendant NGRI, all non-dispositionally instructed juries found in favour of guilt. Evidently, not only does deliberation have a significant impact on verdict decisions, but

dispositional instructions also can help to educate jurors.

**Project Overview**

Researchers have made great gains in understanding individual juror decision-making in insanity cases, but the jury deliberation process remains largely a mystery. The purpose of this dissertation was to examine a potential missing variable (i.e., punishment orientation), and to better understand which phenomena contribute to jury deliberations in an insanity case. To this aim, I analyzed transcribed mock jury deliberations, in which community participants discussed a fictional second-degree murder case involving an NCRMD claim owing to schizophrenia.

**Research questions.**

1. Did jurors rely on insanity myths and express misinformation about the defence? Did other jurors tend to correct this misinformation?
2. Did jurors rely on and express retributive or utilitarian punishment motives?
3. Did jurors discuss the defendant's potential disposition or punishment? Did dispositional instructions elicit certain deliberation content?
4. Did jurors tend to favour the Crown or Defence's narrative? Did alternative narratives arise?
5. Did legal instructions feature in mock jurors' persuasive tactics? If so, how were they mobilized differently for guilty versus NCRMD verdicts?
6. What general strategies did juries rely on? Did they tend to straw poll early in deliberation?
7. Did jurors tend to articulate arguments, or did they appeal to intuition or

emotion?

Study 1 focused on individual juror decision-making, including a Challenge for Cause procedure, while Study 2 situated participant responses in the group context via theory-driven content analysis. Finally, Study 3 examined that content in greater depth, via a qualitative description. Data collection occurred in two phases: an online task and the in-lab deliberation. For efficiency, I describe the full materials and procedure as they apply to all three studies in the Study 1 Method section.

## Study 1

Jurors have notoriously negative attitudes toward the insanity defence, and yet correcting misinformation alone appears insufficient to change verdict decisions (Maeder et al., 2015). It is possible that considering jurors' punishment orientation will help to tease apart different motivations for this negativity. It is further possible that jurors will explicitly identify as against the insanity defence, and thus that a Challenge for Cause procedure might be fruitful in such cases. In Study 1, I analyzed relationships among these individual differences in order to gauge the various personal perspectives that mock jurors bring, outside of group influences.

### Hypotheses

**Challenge for cause.** I expected that endorsing abolition of the insanity defence would be associated with increased likelihood of a guilty verdict, more negative insanity defence attitudes, and permissive utilitarianism and retributivism. Further, indicating that one's opinions about the insanity defence would interfere with the ability to serve on an insanity case (i.e., self-exclusion) was expected to be associated with individual verdict decisions, insanity defence attitudes, and punishment orientation in the same manner. I

acknowledge past social psychological research showing that sometimes people either lack the self-insight necessary to understand their attitudes, or they misreport them (Paulhus & Vazire, 1991). Indeed, in their 2015 study, Schuller and colleagues found that while only 8.3% of participants self-excluded in a race-based challenge for cause, 20.9% of triers (i.e., citizens from the jury pool who evaluate the impartiality of potential jurors) deemed a person partial. Schuller et al. (2009) have also shown no significant differences among self-excluding and self-including responses in a race-related case.

However, I argue that insanity opinions more closely parallel death penalty opinions (i.e., an explicit moral stance). Findings on death qualification indicate that asking jurors to self-exclude tends to disproportionately eliminate those who would always vote against death (as compared to those who would always vote for death), while remaining jurors are prosecution-prone (Neises & Dillehay, 1987). Hence, those with strong objection rather than ambivalence toward or preference for capital punishment are often excluded. It is therefore possible that those with strong objections to the insanity defence would be more apt to self-exclude, but if included in a jury, more likely to vote guilty.

**Punishment orientation.** Following Yamamoto and Maeder (2019), I predicted that permissive retributivism and utilitarianism would be associated with increased likelihood of a guilty verdict. I also expected that prohibitive retributivism and utilitarianism would be associated with decreased likelihood of a guilty verdict, more favourable insanity defence attitudes, and lower likelihood of Challenge for Cause exclusion.

**Insanity defence attitudes.** I predicted that greater strict liability and injustice

and danger attitudes would be associated with greater likelihood of a guilty verdict (Maeder et al., 2015; Skeem et al., 2004), greater permissive retributivism and utilitarianism, lower prohibitive retributivism and utilitarianism, and finally greater likelihood of exclusion via the Challenge for Cause questions.

## Method

**Participants.** Overall, 172 people interacted with the online survey. Of those, 107 completed Phase 1 (i.e., 65 people did not complete the survey.) A total of 24 participants dropped out of the study prior to Phase 2. Remaining participants were 83 (47 men, 34 women, 2 transgender individuals) Canadian jury-eligible community members (i.e., citizens at least 18 years of age with no indictable offences) recruited online via Kijiji, having a mean age of 29 (*SD* = 11.5) and ranging from 18 to 62. The majority of participants (63.9%) identified as White, while 18.1% identified as Black/African-Canadian, 4.8% as Middle Eastern, 2.4% as Aboriginal Canadian/Native Canadian/First Nations, 2.4% as East Indian, 1.2% as Asian, 1.2% as Hispanic/Latino, and 6% as another group.

***A priori sample size.*** A power analysis for a two-tailed Pearson or Point-biserial correlation using G*Power yielded a minimum sample size of 82 for a medium effect size (.30) at $\alpha = .05$, with .80 power. However, due to the exploratory, mixed methods nature of the study, the main rationale for the sample size overall rested on the concept of saturation. While somewhat of a nebulous concept, saturation roughly constitutes reaching the point at which new information has been exhausted; that is, there are sufficient data to be trustworthy (Fusch & Ness, 2015). As Fusch and Ness (2015) articulated, there is no ‘one size fits all’ with respect to minimum raw data quantity, but

instead it is more about properties of the sample. For instance, they cite focus groups (of between six and 12 participants) as one effective means through which to achieve saturation, given the potential diversity of opinions (Fursch & Ness, 2015). The deliberation sessions are akin to focus groups, in that they require several participants to discuss an issue as facilitated by the research assistant, who notably had structured interactions with participants in hopes of minimal interference.

Underscoring the lack of established rules for achieving saturation, Francis and colleagues (2010) outlined four principles that might justify these decisions in theory-driven content analysis. Moreover, they argued that sample size is a practical as well as ethical issue, given that both recruiting too few and too many participants wastes valuable resources. In the first two steps, researchers must *a priori* select an initial analysis sample and a stopping criterion. The initial analysis sample should be based on the minimum sample size needed to satisfy stratification factors (e.g., diversity of age, ethnicity). The stopping criterion dictates the number of additional interviews after new ideas are considered exhausted. The final two principles concern demonstrating data quality through inter-rater reliability of at least two independent coders, and clear reporting of data saturation criteria.

For the current studies, due to inherent limitations in recruitment numbers, we first considered a jury sufficient if there were at least five members (half of the minimum permitted attrition to continue a trial). Notably, we also collected data from multiple juries with the minimum legal requirement of 10 people. Given the large size of the groups, we used six juries as our initial sample size and collected from four extra juries as our stopping criterion. Because of the instruction manipulation, we sought an even

number of juries. Indeed, these criteria yielded a gender balanced sample, with age diversity, which roughly mimicked the racial composition of Canada, albeit with somewhat greater diversity.

***Evidence of saturation.*** Applying all of the aforementioned principles to the current studies reasonably evinces saturation. Most significantly, Fusch and Ness (2015) underscored the need for striking a balance between 'rich' (i.e., quality) and 'thick' (i.e., quantity) data. The paradigm in this dissertation was conducive to rich data. The trial transcript was crafted and piloted so as to be ambiguous in case fact interpretation, and the turnout rate was high (between five and 12 participants). Throughout the deliberations, I monitored the differential interpretation of evidence across new juries. For instance, two different understandings of the defendant's utterance at the time of arrest consistently emerged. Whereas some participants believed the sentence "Don't let them take me" was directed at the officers in hopes to avoid capture by aliens (consistent with a paranoid delusion), others believed it was directed at his mother in hopes to avoid arrest. I also took note of the overall narrative and the verdict breakdowns.

Granted, as Fusch and Ness (2015, p. 1411) cautioned, in such studies, the investigator is to some extent "the data collection instrument," and hence it is necessary to recognize one's personal "lens" when assessing saturation. Consistent with Devine's (2012) Director's Cut Model of JDM, I observed repetition of a handful of main narratives (e.g., doubting versus accepting the prosecution's story), which the jury system is itself designed to elicit. There was a discernable pattern at roughly Jury 5, and we ended data collection after twice as many juries. Additionally, the research assistants did not provide hints or opinions even when probed by participants, but instead only re-read

instructions verbatim. In sum, while it is difficult to articulate a steadfast sample size rule for these data, I maintain that 10 juries (comprised of five to 12 participants each) are sufficient for this dissertation. The sample size is powered for the purpose of the individual-level analyses, and there is some evidence of saturation given the emergence of consistent narratives and balanced verdict splits.

**Materials: Phase 1**

***Insanity defence attitudes.*** Participants completed the 19 items of the Insanity Defense Attitudes-Revised Scale (IDA-R; Skeem et al., 2004, see Appendix A) adapted to a Canadian context, which comprises two latent factors (injustice and danger, strict liability). The Strict Liability scale pertains to the extent that a person believes that mental disorder is irrelevant to criminal responsibility (e.g., "I believe that people should be held responsible for their actions no matter what their mental condition") and showed strong internal consistency ($\alpha = .85$). The Injustice and Danger scale pertains to fears about misuse of the defence and the potential threat to public safety (e.g., "As a last resort, defence attorneys will encourage their clients to act strangely and lie through their teeth to appear mentally ill") and showed strong internal consistency ($\alpha = .86$). Participants rated their agreement on a 7-point Likert-type scale (where *1 = strongly disagree*, and *7 = strongly agree*).

***Punishment orientation.*** Participants also completed the 17 items of the Punishment Orientation Questionnaire (POQ; Yamamoto & Maeder, 2019, see Appendix A), which comprises four scales that measure the principles people rely on when thinking about appropriate punishment in the criminal justice system. The Prohibitive Utilitarian scale measures the extent to which participants believe punishment should be goal-

oriented and benefit society (e.g., "Punishment should be about looking forward to improve society, not backward to address the criminal's misdeeds"); the scale showed strong internal consistency ($\alpha = .82$). The Permissive Utilitarian scale measures the extent to which participants are willing to give strict punishment to the aim of deterrence (e.g., "an overly harsh punishment may be necessary to prevent others from committing the same crime"); the scale showed strong internal consistency ($\alpha = .82$). The Prohibitive Retributive scale captures aversion to the risks of punishment (e.g., "It is better to let 10 guilty criminals go free than to punish 1 innocent person"); the scale showed strong internal consistency ($\alpha = .81$). Finally, the Permissive Retributive scale captures blame of the criminal label and desires for retribution ("Criminals are bad people and get what is coming to them"); the scale showed strong internal consistency ($\alpha = .84$). Participants rated their agreement on a 5-point Likert scale (where *1 = strongly disagree*, and *5 = strongly agree*).

***Challenge for cause.*** We asked participants two questions related to their ability to serve on a jury in an insanity case. First, we asked participants (Yes/No) whether they "believe that the insanity defence should be abolished." Second, we asked participants (Yes/No): "would your ability to judge the evidence in this case without bias, prejudice, or partiality be affected by the fact that the accused is pleading Not Criminally Responsible on Account of Mental Disorder?" Instructions indicated that responses would not affect eligibility to participate in the remainder of the study. Implications for this decision are considered in the Study 1 Discussion section.

***Demographics survey.*** Participants completed a demographics survey, which included race, gender, occupation, and level of education. They were also asked to

provide their religious and political affiliations (if any), and whether they personally know someone with a mental disorder. Finally, participants were asked to indicate where their political beliefs fell on a sliding liberal to conservative scale.

**Materials: Phase 2**

***Pre-trial instructions.*** Participants heard one page of model jury instructions adapted from those provided by the National Judicial Institute (2014) about the essential elements of the charge and requirements for NCRMD, as well as the burden of proof.

***Trial transcript.*** I created an approximately 8-page trial transcript (see Appendix B) loosely based on *Clark v. Arizona* (2006), which describes a second-degree murder charge against a man who stabbed his roommate. The transcript begins with opening statements from the Crown and Defence. The Crown alleges that the accused is a violent man who snapped in response to a heated argument. The police officer who arrested the defendant serves as a Crown witness and provides evidence that the defendant was attempting to flee with the victim's wallet. The Defence alleges that the accused had paranoid schizophrenia at the time of the crime and specifically had Capgras delusion. A psychiatrist (whose gender was left ambiguous) testifies to this effect. The psychiatrist describes the diagnostic criteria for the defendant's disorder and gives an explanation for the lack of clear history of mental disorder. The trial ends with closing statements from the Crown followed by the Defence.

***Individual verdict decision.*** Each participant filled out an individual verdict form after reading the trial transcript. Participants selected from guilty, not guilty, or not criminally responsible on account of mental disorder.

***Pre-deliberation instructions.*** Participants heard two pages of instructions about

the criteria for NCRMD and how to decide whether the defence meets those requirements. These instructions also reiterate special rules on the burden of proof and reasonable doubt. Half of the juries heard an additional paragraph of dispositional instructions as follows:

> Your verdict must be based solely on your consideration of all the evidence, the submissions of counsel, and the law as I have explained it to you and must not be influenced by a consideration of the consequences of a special verdict of not criminally responsible. However, you should know that our law provides that persons who are found not criminally responsible by reason of mental disorder will not be set free if they pose a significant threat to public safety. (National Judicial Institute, 2014)

Finally, participants were instructed on logistics of the deliberation (e.g., selecting a foreperson, unanimity requirements; National Judicial Institute, 2014). Appendix B displays these instructions.

***Jury verdict decision.*** The foreperson was instructed to complete a verdict form, selecting from guilty, not guilty, NCRMD, or unable to reach a verdict.

**Procedure: Phase 1.** Three to four days prior to the deliberation, participants followed a link from a recruitment notice on Kijiji. After passing juror eligibility screening, participants selected the appropriate time slot for the coming Saturday (and were told to check back the following week for alternative sessions). They were then directed to the Phase 1 informed consent form, followed by the (counterbalanced) IDA-R and POQ. Participants then completed the Challenge for Cause questions followed by a brief demographics survey. After being directed to a new survey, participants either

entered an email address and received further instructions for the deliberation phase, or they withdrew from the study.

**Procedure: Phase 2.** We sent participants a reminder email the afternoon prior to Phase 2. Participants were seated in order of arrival (i.e., the first to arrive was assigned as Juror #1, which was displayed on the table and above the seat). After all participants arrived (or up to 10 minutes after the official start time) the research assistant read welcome instructions outlining the risks of the study and procedures for withdrawing. Once participants completed informed consent, the research assistant read the pre-trial instructions (about 5 minutes) and handed out the trial transcripts with individual verdict forms, for which participants were given about 15 minutes. While juror note-taking is at the discretion of individual judges, participants were not permitted to do so in the current study. Some legal practitioners have expressed concerns that note-taking can bias juror decisions and jury deliberations, but early experiments have shown that note-taking does not appear to aid comprehension nor hinder juror performance (Devine, 2012; Penrod & Heuer, 1997; 1998). We therefore opted for the more simplistic procedure, given the number of variables that note-taking potentially introduces.

After all materials were collected, the research assistant read the pre-deliberation instructions and provided the jury verdict form. Sommers (2006) used a limit of 60 minutes and found that on average deliberations in a sexual assault case ranged from approximately 38 to 50 minutes. Given practical limitations to coding time and funds, participants were told that the deliberation would last no longer than 45 minutes; a clock was displayed on a screen. The research assistant then left the room for the duration of the deliberation and waited in a smaller lab office next door to observe participants

through a two-way mirror and audio-visual system. Participants could wave at the two-way mirror any time for assistance or once they reached a verdict. Juries close to the 45-minute mark were given a 5-minute warning. The research assistant was instructed not to provide any further information about the case, except to re-read passages of the instructions if questioned. Once the deliberation finished, participants completed the post-deliberation questionnaire. They were then debriefed and compensated with $40 for their time.

## Study 1 Results

### Attitudinal Variables

First, I assessed bivariate relationships among the attitudinal variables[1]. Table 1 displays descriptive statistics for all continuous measures. Table 2 shows the correlations.

**Table 1** ***Descriptive statistics for attitudinal variables and verdict confidence.***

| | N | Mean (SD) |
|---|---|---|
| Strict Liability | 82 | 3.27 (1.11) |
| Injustice & Danger | 82 | 3.77 (1.12) |
| Permissive Retributive | 80 | 2.91 (.92) |
| Permissive Utilitarian | 80 | 2.73 (.96) |
| Prohibitive Retributive | 80 | 3.47 (.85) |
| Prohibitive Utilitarian | 80 | 3.79 (.82) |
| Verdict Confidence | 82 | 7.23 (1.93) |

Note: Scales with missing items featured person mean imputation.

[1] The POQ and IDA-R were counterbalanced, but there were no significant differences in reported analyses on this basis.

Table 2 *Bivariate relationships among attitudinal measures.*

| | Strict Liability | Injustice & Danger | Permissive Retributive | Permissive Utilitarian | Prohibitive Retributive | Prohibitive Utilitarian |
|---|---|---|---|---|---|---|
| Strict Liability | 1 | | | | | |
| Injustice & Danger | .49** | 1 | | | | |
| Permissive Retributive | .46** | .42** | 1 | | | |
| Permissive Utilitarian | .36** | .41** | .70** | 1 | | |
| Prohibitive Retributive | -.28* | -.31* | -.30* | -.39** | 1 | |
| Prohibitive Utilitarian | -.36** | -.12 | -.45** | -.22* | .24* | 1 |

** $p \leq .001$
* $p \leq .05$

As anticipated, the IDA-R dimensions shared a moderate positive relationship, and the permissive dimensions of the POQ showed weak negative relationships with the prohibitive dimensions. As in previous studies (Yamamoto & Maeder, 2019) the permissive retributive and utilitarian dimensions were strongly positively correlated.

In line with expectations, the POQ dimensions showed moderate relationships with the IDA-R dimensions. It was conceivable that strict liability would be more strongly related to retributivism, and injustice and danger more strongly related to utilitarianism (Maeder et al., 2015; Skeem et al., 2004). However, permissive utilitarianism and permissive retributivism showed moderate positive relationships with both dimensions. Prohibitive retributivism showed a weak negative relationship with both IDA-R dimensions. Prohibitive utilitarianism showed a weak negative relationship with strict liability only, having no significant linear association with injustice and danger.

Those higher in the tendency to focus on the positive societal impact of punishment were less likely to believe that mental disorder is irrelevant to a crime, but no less likely to believe the insanity defence is misused or threatens public safety.

**Criterion Variables**

Next, I assessed the relationships between the attitudinal measures and criterion variables: belief that the insanity defence should be abolished, self-exclusion in Challenge for Cause, and pre-deliberation (i.e., individual) dichotomous verdict decision. Additionally, a continuous verdict variable (the multiplicative product of guilty versus NCRMD and verdict confidence) was included. The majority of jurors individually rendered a guilty verdict ($n$ = 50, 61%) prior to deliberation, whereas 32 (39%) chose NCRMD. The majority of jurors did not self-exclude from serving on the jury (i.e., responded "no", $n$ = 67, 81.7%), whereas a minority self-excluded ($n$ = 15, 18.3%). Likewise, the majority did not endorse abolition of the insanity defence ($n$ = 71, 86.6%), whereas a minority endorsed abolition ($n$ = 11, 13.4%).

Table 3 *Bivariate relationships between attitudinal measures and criterion variables.*

| | Abolition | Challenge | Individual Dichotomous Verdict | Individual Continuous Verdict |
|---|---|---|---|---|
| Strict Liability | .42** | .19 | .23* | 0.28* |
| Injustice & Danger | .47** | .18 | .25* | 0.32* |
| Permissive Retributive | .29* | -.03 | .24 | 0.30* |
| Permissive Utilitarian | .37** | -.11 | .16 | 0.22* |
| Prohibitive Retributive | -.24* | .03 | .00 | 0.00 |
| Prohibitive Utilitarian | -.13 | .08 | -.08 | -0.12 |

$N$ = 82 ** $p \leq .001$ * $p \leq .05$

Table 3 displays the bivariate relationships between the attitudinal measures and

the criterion variables. Interestingly, the Challenge for Cause question did not evince a significant relationship with any of the attitudinal measures. Endorsement of abolition, however, was significantly related to all measures save for prohibitive utilitarianism. Individual dichotomous verdict related only to the dimensions of the IDA-R. Permissive retributivism and utilitarianism shared a significant positive relationship with continuous verdict, such that those higher on the traits were more confident in a guilty verdict, and those lower on the traits were more confident in an NCRMD verdict.

Additionally, political orientation was significantly related to continuous verdict, $r$ (78) = .32, $p$ = .005, such that higher identification with conservatism was associated with greater confidence in a guilty verdict, or greater identification with liberalism was associated with greater confidence in an NCRMD verdict. Political orientation was also significantly related to permissive retributivism, $r$ (76) = .29, $p$ = .01, such that higher identification with conservatism was associated with greater permissive retributivism, or greater identification with liberalism was associated with lower permissive retributivism.

Table 4 *Contingency tables for pre-deliberation verdict and other variables of interest.*

| | | | Individual verdict | |
|---|---|---|---|---|
| | | | NCRMD | Guilty |
| Do you believe that the insanity defence should be abolished?[a] | Yes | Count | 1(9.1%) | 10(90.9%) |
| | | Adjusted Residual | -2.2* | 2.2* |
| | No | Count | 31(43.7%) | 40(56.3%) |
| | | Adjusted Residual | 2.2* | -2.2* |
| Challenge for Cause response | Yes | Count | 5(33.3%) | 10(66.7%) |
| | | Adjusted Residual | -.5 | .5 |
| | No | Count | 27(40.3%) | 40(59.7%) |
| | | Adjusted Residual | .5 | -.5 |

*Cell significantly contributed to chi-square statistic (i.e., observed count significantly differed from expected count).
a. Contains one cell with fewer than five observed counts.

Table 4 displays contingency tables for analyses involving individual verdicts. I report observed counts, percentages, and adjusted standardized residuals for each cell. Adjusted standardized residuals are a metric of contribution to the chi-square statistic, flagging cells in which the observed count differed significantly from the expected count (Beasley & Schumacker, 1995). The critical cut-off to judge statistical significance was an absolute value of 1.96. Self-exclusion did not show a significant relationship with individual verdict, $\chi^2(1) = 0.25$, $p = .617$, $v = .06$. However, endorsement of abolishing the insanity defence shared a significant relationship with individual verdict, $\chi^2(1) = 4.74$, $p = .029$, $v = .24$. Among those who endorsed abolition there was a higher than expected guilty verdict count. Of note, abolition was also significantly positively correlated with continuous verdict, $r\ (82) = .27$, $p = .01$. I then examined whether tendency to self-exclude in response to the Challenge for Cause question was significantly related to belief that the insanity defence should be abolished. A chi-square test, $\chi^2(1) = 0.69$, $p = .41$, $v = .09$, did not reveal any such contingency. Finally, I examined a contingency table to test for a relationship between the criterion variables, which was non-significant. (see Table 5).

Table 5 *Non-significant relationship between self-exclusion and abolition.*

| | | | Challenge for Cause Response | |
|---|---|---|---|---|
| | | | Yes | No |
| Do you believe that the insanity defence should be abolished? | Yes | Count | 3(27.3%) | 8(72.7%) |
| | | Adjusted Residual | .8 | -.8 |
| | No | Count | 12(16.9%) | 59(83.1%) |
| | | Adjusted Residual | -.8 | .8 |

## Study 1 Discussion

The purpose of Study 1 was to assess the role of punishment orientation and insanity defence attitudes in individual juror decisions. Further, Study 1 addressed the utility of a Challenge for Cause question – hypothetical self-exclusion from the jury due to bias – in predicting verdict decisions. Participants also indicated whether they believed the insanity defence should be abolished.

The Challenge for Cause question was not a useful gauge of individual verdict decisions. Challenge for Cause also shared no significant contingency with endorsement of abolition. This is perhaps unsurprising; researchers have shown that jurors are by and large unable to identify or control for their own biases (see Miller & Bornstein, 2017). The criteria for Challenge for Cause questions are decidedly rigid (allowing little clarification) and imprecise. The single item depends on a) jurors' accurate perceptions that their beliefs are biased, and b) an evaluation of the impact of that bias. Moreover, a tradition of social psychology highlights the many barriers to accurate self-knowledge (Brown & Dutton, 1995; Nisbett & Wilson, 1977; Taylor & Brown, 1988). On the other hand, a question regarding endorsement of abolition only requires the participant to evaluate their stance on a particular issue. Indeed, endorsement of abolishing the insanity defence was significantly related to verdicts; those who endorsed abolition were more likely to vote guilty prior to deliberation. It bears mentioning that the issues with Challenge for Cause are likely to be exacerbated in the context of a real case, in which jurors are asked questions in person and in front of others. Research suggests that people are more disinhibited online (Joinson, 1999), which could theoretically alleviate social desirability concerns. However, it is also possible that participants put less effort into

considering the question of self-exclusion than they might in a case having higher stakes. At any rate, this finding supports that of Schuller et al. (2009), who observed no significant differences in verdict decision as a function of self-exclusion (in a race-related case). Of course, these results are preliminary and further testing is needed to rule out the usefulness of Challenge for Cause in insanity cases.

In line with speculation from Skeem et al. (2004) and Maeder et al. (2015), the dimensions of the IDA-R did share some relationship with punishment orientation. As Breheney et al. (2007) remarked, jurors might be uncomfortable with the notion that those found insane are not punishable in the traditional sense. Again, in line with previous findings (Maeder et al., 2015), the dimensions of the IDA-R were significantly related to individual dichotomous verdict decisions. Permissive utilitarianism and retributivism appeared unrelated to dichotomous verdict but showed significant relationships with continuous verdict. Participants higher on these traits showed higher confidence in a guilty verdict, and those lower on the traits showed higher confidence in an NCRMD verdict. One possible explanation for this finding is that people with stronger convictions about their verdict decision were more likely to have strongly developed beliefs about punishment.

Among the dimensions of the POQ, prohibitive utilitarianism yielded the most surprising results. It did not relate strongly to the other attitudinal or criterion variables. While all other attitudinal measures related significantly to abolition endorsement, again prohibitive utilitarianism did not. It is possible that this represents substantive information about the prohibitive utilitarian trait. Perhaps the tendency to focus on the positive societal effects of punishment does not weigh into the issue of abolition.

However, Yamamoto and Maeder (2019) indicated that this dimension of the POQ might simply be weaker. The scale showed some unstable psychometric properties, and it is yet unclear theoretically what comprises these attitudes. Attention to whether jurors discussed any such principles during deliberation might shed light on potential reasons for this finding.

There are some notable limitations to Study 1. First, the small sample size prohibits more sophisticated quantitative analyses than those reported. These data still provide valuable impressions of the linear relationships among attitudes and pre-deliberation decisions. Granted, participants only completed a handful of attitudinal measures, and so the data capture an incomplete picture. Second, the case appears to be somewhat guilt leaning, which may itself contribute to the relationship with punishment orientation or insanity defence attitudes. The POQ is a relatively new measure, and so it is unclear to what extent permissive and prohibitive retributivism and utilitarianism are context dependent. For instance, it could be that particularly heinous cases render permissive retributive concerns more central. However, Yamamoto and Maeder (2019) reported that presenting the POQ before versus after a death penalty case did not significantly influence findings, suggesting that the measure could be context resistant. The deliberation analyses help to underscore relevant idiosyncrasies of the trial transcript by showing what narratives participants created from the evidence. Finally, the decision to inform participants that their response to the Challenge for Cause question would not influence eligibility might have impacted on results. This choice somewhat detracts from the ecological validity of the study, because participation would be at stake in real life. However, this strategy was intended to prevent acquiescence, given that participants

knew they would not receive payment if they did not participate in Phase 2. Further, real jurors might be motivated to avoid rather than participate in jury duty, and so not informing mock jurors on the consequences of their decision could also be influential.

In brief, it appears that Challenge for Cause is insufficient to predict individual juror decisions in this context. Instead, endorsement of abolition might be a more useful gauge of potential verdict decisions. Further, punishment orientation indeed seems to play some role in mock jurors' beliefs about the insanity. The next step was to ascertain what topics mock jurors actually leveraged in attempts to publicly defend their positions.

## Study 2: Overview

Study 2 concerned the examination of mock jury deliberation sessions. Hsieh and Shannon (2005) described the method of *directed content analysis* as a means to further investigate established phenomena, consistent with the aims of this dissertation. To do so, the researcher begins with an initial set of coding categories arising from previous findings, which are then operationalized. The researcher may then assess a specified unit of analysis and code each utterance in line with the pre-established taxonomy. Likewise, in Study 2 I employed a deductive method, given that previous findings on the insanity defence guided a priori creation of a coding manual. However, I conducted a pilot test with student participants to allow for unanticipated coding categories. In the following sections, I describe creation of this coding manual and the procedure for conducting analyses. Of note, deliberation videos were transcribed by several lab volunteers. Transcribers were asked to indicate which juror was speaking, which acted as data sources.

### Hypotheses

**Hypothesis 1a: Crown position-taking.** Researchers have theorized that some jurors dislike the insanity defence because it implies the defendant is not punishable (Breheney et al., 2007; Skeem et al., 2004). Consequently, I predicted that more permissive punishment orientation would be associated with a higher frequency of Crown position-taking (i.e., expressing the opinion that the defendant was likely guilty).

**Hypothesis 1b: Defence position-taking.** I predicted that lower permissive punishment orientation would be associated with a higher frequency of Defence position-taking (i.e., expressing the opinion that either prong of NCRMD might be met, or that the defendant was NCRMD).

**Hypothesis 2: Legal instructions.** Because the burden is on the defendant to prove that mental disorder precludes criminal responsibility, I hypothesized that those who want punishment would rely on this standard. I predicted that more permissive punishment orientation would be associated with a higher frequency of references to legal instructions (i.e., reasonable doubt, the burden of proof, the legal requirements of NCRMD).

**Hypothesis 3: Psychiatrist credibility.** I predicted that more permissive punishment orientation would be associated with a lower frequency of references to the Psychiatrist's credibility or authority.

**Hypothesis 4: Defendant disposition.** I hypothesized that punishment-prone mock jurors would be more preoccupied with the defendant's disposition. Thus, I predicted that more permissive punishment orientation would be associated with a higher frequency of references to the defendant's ultimate disposition (i.e., what would happen to the defendant following trial).

**Coding**

Coders used NVivo, a qualitative coding software. They selected from a list of codes (see Appendix C). To create this codebook, I relied on principles arising from Skeem et al.'s (2004) IDA-R and Yamamoto and Maeder's (2019) POQ. The pilot deliberation yielded several key pieces of evidence, which the jurors discussed at length. Finally, the codebook featured some emotion-based constructs consistent with the rationalist and intuitionist accounts of morality, which were intended to aid in showing the extent to which jurors have either proclivity.

**Segmenting**

Roughly mirroring Greene, Hayman, and Motyl (2008), coders examined uninterrupted utterances on a single topic (i.e., "idea units", p. 208), with sentences as the rough grain size (Chi, 1997). Implications for this choice of granularity are explored in the General Discussion section. Because people speak less formally in comparison to written communication, punctuation was only one potential marker for a coding unit; the utterance had to express a coherent thought. These units ranged from two-word ideas (e.g., "I agree") to several word run-on sentences (e.g., "And I think he knew that like it was uh it was yea, y'know, yea like, I i-illegal yes, but morally –"). Coders were conservative in applying labels to passages. A slight degree of ambiguity resulted in a label of "other"; for example, it might have been unclear with what concept a participant was agreeing. Codes were only applied where there was substantive content. The scheme was exhaustive for each unit (i.e., only one code was applied to each sentence) with the exception of scope, which was later applied to Crown and Defence evidence related utterances.

**Inter-rater reliability**

As Mouter and Vonk Noordegraaf (2012) remarked, transparency and replicability are the chief aims of reliability assessments; hence, at least two raters are needed for the same content. Two independent coders were trained to assess each individual utterance for the categories listed in the codebook. I used Cohen's kappa as a metric of reliability and resolved disagreements through discussion. Cohen's Kappa confers the benefit of accounting for chance, and it is seamlessly integrated with the NVivo software. Convention suggests that a kappa between .40 and .75 is fair, and anything above is excellent (Cohen, 1960). Coders were blind to the type of instruction provided (with disposition vs. without disposition).

Following Sommers (2006), two coders assessed 20% of the juries, and one coder assessed the remaining juries. To assess inter-rater reliability, I used NVivo's Coding Comparison Query tool, basing calculations on sentences. Notably, some codes were used infrequently, which yielded perfect agreement when they were not present, but did little to illustrate coders' ability to detect that content. For instance, coders had 100% agreement regarding instances of "suspension of disbelief", because only one juror indicated that the transcript was likely fictional. Notably, several of the disagreements amounted to coder disagreement about only one utterance, solved easily through discussion.

Overall, the range of Kappas demonstrated fair reliability (.40 to 1.00, Cohen, 1960). Table 6 displays the Kappa coefficients. Several values were at the lower end of the conventionally acceptable Kappa range. Low base rates can decrease Kappa even in cases of high agreement (Xu & Lorber, 2014) and so again the infrequency of some codes

warrants caution in interpreting these values. However, after an initial round of coding, it became clear that direct references to utilitarian and retributive punishment principles, as well as to specific insanity defence myths were simply too rare for inter-rater agreement to be quantified in a trustworthy manner (see Study 2 Discussion section for implications). The coding scheme was therefore amended to flag participants' expression of general beliefs about the insanity defence and mental disorders, and to mentions of the defendant's eventual disposition. Likewise, some of the categories were not amenable to the type of manifest content coding applied in Study 2 (i.e., emotion-related codes, moral conviction). As a solution, alternative strategies relying on computer automation and/or qualitative methods were later employed in an exploratory third study.

Table 6 *Interrater reliability Kappa coefficients*

| Code | Jury A | Jury F |
|---|---|---|
| IDA's | .52 | .55 |
| Disposition | .25 | .65 |
| Legal Instructions | .61 | .58 |
| Crown Evidence | .91 | .54 |
| Defence Evidence | .63 | .68 |
| Scope | .58 | .56 |
| Anecdotes | 1.00 | .77 |
| Suspension of Disbelief | 1.00 | .61 |
| Position-taking | .73 | .63 |

## Study 2: Results

### General Characteristics

Tables 7 and 8 display an overall summary of the features of each jury (group features and demographics respectively). Deliberations ranged from 1.7 minutes to 47 minutes ($M$ = 34.5, $SD$ = 13.5). Jury 10 had the shortest discussion time by a substantial margin, and so it was considered an outlier. Without Jury 10, deliberations ranged from 28 minutes to 47 minutes ($M$ = 38.2, $SD$ = 7.5). Juries ranged in size from 5 to 12 people. Seven juries elected a foreperson at the outset of deliberation (i.e., within 10 to 150 utterances), while three juries elected a foreperson at the end of deliberation (i.e., when filling out the verdict form).

Among the three juries that ultimately voted guilty, 5 participants rendered individual NCRMD verdicts. Only one of those juries featured a lone NCRMD vote. Table 9 displays the rate of major topics for each of the 10 juries. A total of 4871 utterances were coded, with a mean of 493.5 per jury ($SD$ = 122.4). Given that the data had several zero counts, means and standard deviations were not a useful metric of central tendency (see Table 10 for median, mode, and range for each code).

Table 7 *Group characteristics broken down by discussion group.*

| Group | Outcome | Proportion Guilty Verdicts | Size | Straw Poll Timing* | Time (Minutes) |
|---|---|---|---|---|---|
| Jury 1 (A) | Hung | .56 | 9 | 326 (549) | 41.00 |
| Jury 2 (B) | Guilty | .67 | 6 | 325 (327) | 28.00 |
| Jury 3 (C) | Hung | .33 | 9 | Informal Sequential (618) | 47.00 |
| Jury 4 (D) | Guilty | .82 | 11 | 1 (372) | 28.00 |
| Jury 5 (E) | NCRMD | .25 | 8 | Informal Sequential** (557) | 42.00 |
| Jury 6 (F) | NCRMD | .50 | 6 | Formal sequential (637) | 36.00 |
| Jury 7 (G) | Hung | .50 | 6 | 391 (730) | 32.00 |
| Jury 8 (H) | Hung | .75 | 12 | 14 (383) | 47.00 |
| Jury 9 (I) | Hung | .67 | 9 | 29 (451) | 43.00 |
| Jury 10 (J) | Guilty | .86 | 7 | 26 (31) | 1.72 |

Notes: "Informal sequential" indicates that jurors provided a verdict along with a rationale and/or informally expressed positions in turn. "Formal sequential" indicates that jurors confirmed a position one after another.
*Utterance number at which a formal poll was initiated and completed, followed by total number of utterances in brackets.
**Poll vetoed by a juror.

Table 8 *Demographics broken down by discussion group.*

| Group | Age | Gender | | | Racial Composition | | Know Person with Mental Disorder | |
|---|---|---|---|---|---|---|---|---|
| | Mean (SD) | Man | Woman | Trans | White | Another Race | Yes | No |
| Jury 1 (A)* | 30.2 (9.7) | 6 (66.7%) | 3 (33.3%) | 0 | 6 (66.7%) | 3 (33.3%) | 4 (44.4%) | 5 (55.6%) |
| Jury 2 (B)** | 25.0 (5.1) | 5 (83.3%) | 1 (16.7%) | 0 | 2 (33.3%) | 4 (66.7%) | 4 (66.7%) | 2 (33.3%) |
| Jury 3 (C)* | 40.4 (16.8) | 4 (44.4%) | 5 (55.6%) | 0 | 7 (77.8%) | 2 (22.2%) | 5 (55.6%) | 4 (44.4%) |
| Jury 4 (D)** | 26.6 (9.4) | 7 (63.6%) | 3 (27.3%) | 1 (9.1%) | 5 (45.5%) | 6 (54.5%) | 7 (63.5%) | 4 (36.4%) |
| Jury 5 (E)*** | 24.8 (9.1) | 6 (75.0%) | 2 (25.0%) | 0 | 4 (50.0%) | 4 (50.0%) | 5 (62.5%) | 3 (37.5%) |
| Jury 6 (F)*** | 30.0 (14.0) | 3 (50.0%) | 3 (50.0%) | 0 | 5 (83.3%) | 1 (16.7%) | 4 (66.7%) | 2 (33.3%) |
| Jury 7 (G)* | 23.8 (7.2) | 4 (66.7%) | 2 (33.3%) | 0 | 3 (50.5%) | 3 (50.5%) | 4 (66.7%) | 2 (33.3%) |
| Jury 8 (H)* | 22.7 (4.6) | 5 (41.7%) | 7 (58.3%) | 0 | 9 (75.5%) | 3 (24.5%) | 7 (58.3%) | 5 (41.7%) |
| Jury 9 (I)* | 26.9 (8.1) | 3 (33.3%) | 5 (55.6%) | 1 (11.1%) | 7 (77.8%) | 2 (22.2) | 9 (100%) | 0 (0%) |
| Jury 10 (J)** | 40.0 (12.7) | 4 (57.1%) | 3 (42.9%) | 0 | 5 (71.4%) | 2 (28.6%) | 4 (57.1%) | 3 (42.9%) |

*Hung **Guilty ***NCRMD

Table 9 *Rate of major topics (per total utterance count).*

| Group | Outcome | Crown | | Defence | | Disposition | | Mental Disorder | | Legal Instructions | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Count | Proportion | Count | Proportion | Count | Proportion | Count | Proportion | Count | Proportion |
| Jury 1 | Hung | 22 | .04 | 17 | .03 | 10 | .02 | 46 | .09 | 68 | .13 |
| Jury 2 | Guilty | 21 | .06 | 2 | .01 | 5 | .02 | 15 | .05 | 13 | .04 |
| Jury 3 | Hung | 15 | .02 | 15 | .02 | 14 | .02 | 63 | .10 | 48 | .08 |
| Jury 4 | Guilty | 24 | .06 | 5 | .01 | 11 | .03 | 37 | .10 | 15 | .04 |
| Jury 5 | NCRMD | 3 | .01 | 24 | .04 | 21 | .04 | 25 | .04 | 80 | .14 |
| Jury 6 | NCRMD | 14 | .02 | 20 | .03 | 56 | .09 | 22 | .03 | 44 | .07 |
| Jury 7 | Hung | 16 | .02 | 12 | .02 | 12 | .02 | 37 | .05 | 63 | .09 |
| Jury 8 | Hung | 17 | .04 | 11 | .03 | 24 | .06 | 45 | .12 | 39 | .10 |
| Jury 9 | Hung | 17 | .04 | 9 | .02 | 28 | .06 | 38 | .08 | 47 | .10 |
| Jury 10 | Guilty | 7 | .23 | 0 | .00 | 0 | .00 | 0 | .00 | 3 | .10 |

Table 10 *Utterance content per juror.*

| Code | Median | Mode | Range |
|---|---|---|---|
| Crown Position-taking | 1 | 1 | 13 |
| Defence Position-taking | 0 | 0 | 13 |
| Belief about insanity or mental disorder | 2 | 0 | 33 |
| Legal Instructions | 3 | 0 | 27 |
| Disposition | 1 | 0 | 19 |
| Psychiatrist | 1 | 0 | 7 |
| Anecdotes | 0 | 0 | 14 |
| Suspension of disbelief | 0 | 0 | 7 |
| Other | 10 | 1 | 142 |
| Total utterances | 43 | 2 | 251 |

**Turn-Taking**

To gauge patterns of participation in discussion, I examined the number of continuations (i.e., uninterrupted strings of utterances on different ideas). To do so, I used SPSS's sequential case processing and lag function. As Table 11 shows, the majority of utterances occurred as exchanges. The longest uninterrupted chain of speech was 21 consecutive units.

Table 11 *Counts of uninterrupted utterance chains on single ideas.*

| Number of Consecutive Idea Units | Frequency | Percentage |
|---|---|---|
| 1 | 3542 | 72.7 |
| 2 | 635 | 13.0 |
| 3 | 286 | 5.9 |
| 4 | 149 | 3.1 |
| 5 | 87 | 1.8 |
| 6 | 53 | 1.1 |
| 7 | 35 | .7 |
| 8 | 18 | .4 |
| 9 | 13 | .3 |
| 10 | 12 | .2 |
| 11 | 9 | .2 |
| 12 | 7 | .1 |
| 13 | 7 | .1 |
| 14 | 5 | .1 |
| 15 | 4 | .1 |
| 16 | 3 | .1 |
| 17 | 2 | .0 |
| 18 | 1 | .0 |
| 19 | 1 | .0 |
| 20 | 1 | .0 |
| 21 | 1 | .0 |

**Main Analyses**

For each hypothesis, I tested a separate hierarchical linear model using HLM Software 7 (Raudenbush, Bryk, & Congdon, 2017). Given that the dependent variables were based on counts rather than continuous measures and had several zero counts (i.e.,

were positively skewed, but not suitable for regular transformations), the data did not meet the assumptions of ordinary linear regression (Gardner, Mulvey, & Shaw, 1995). I therefore specified a Poisson distribution with over-dispersion (Raudenbush, Bryk, & Congdon, 2017). Further, because jurors varied in number of utterances, I included total individual utterance count as an exposure variable, which accounted for different "chances" for observation of each code category. First, I examined the null model to ascertain whether the jury that a participant was in significantly contributed to the variance in the dependent variable. That is, I ran a model without any independent variables. Second, I added grand-mean-centred punishment orientation score[2] (where higher scores denote more permissive orientation) as a level 1 predictor and executed a random-intercepts only model. Results reported represent population average models.

**Hypothesis 1a: Crown Position-Taking**

In Hypothesis 1a, I predicted that more permissive punishment orientation would be associated with a higher frequency of Crown position-taking (i.e., expressing the opinion that the defendant was likely guilty). Using Crown position-taking frequency as the dependent variable, the null model was significant, $\chi^2(9) = 26.87, p = .002$, demonstrating that level 2 grouping significantly contributed to the variation in Crown position-taking frequency. Punishment orientation did not significantly predict Crown position-taking ($B = 0.38, SE = 0.20, p = .060$).

[2] Due to the correlations between POQ and IDA-R scores, as well as between the POQ dimensions, I included only a combined punishment orientation score as a predictor, considering that the role of punishment was a central thesis.

**Hypothesis 1b: Defence Position-Taking**

In Hypothesis 1b, I predicted that lower permissive punishment orientation would be associated with higher frequency of Defence position-taking (i.e., expressing the opinion that either prong of NCRMD might be met, or that the defendant was NCRMD). Using Defence position-taking frequency as the dependent variable, the null model was non-significant, $\chi^2(9) = 9.60, p = .384$, demonstrating that level 2 grouping did not significantly contributed to the variation in Defence position-taking frequency. There was a significant effect of punishment orientation on Defence position-taking frequency, $B = -0.56, SE = 0.17, p = .002$, such that those with a higher permissive orientation had a lower frequency of Defence position-taking utterances.

**Hypothesis 2: Legal Instructions**

In Hypothesis 2, I predicted that more permissive punishment orientation would be associated with a higher frequency of references to legal instructions (i.e., reasonable doubt, the burden of proof, the legal requirements of NCRMD). Using legal instruction-related utterance frequency as the dependent variable, the null model was significant, $\chi^2(9) = 24.87, p = .004$. The effect of punishment orientation on legal instruction utterances was non-significant, $B = -0.20, SE = 0.14, p = 0.155$.

**Hypothesis 3: Psychiatrist Credibility**

In Hypothesis 3, I predicted that more permissive punishment orientation would be associated with a lower frequency of references to the Psychiatrist's credibility or authority. Using Psychiatrist credibility-related utterance frequency as the dependent variable, the null model was significant, $\chi^2(9) = 25.70, p = .003$**.** There was a significant effect of punishment orientation, $B = -0.42, SE = 0.18, p = .024$, such that more

permissive orientation was associated with a decreased likelihood of psychiatrist credibility related utterances.

**Hypothesis 4: Disposition.**

In Hypothesis 4, I predicted that more permissive punishment orientation would be associated with a higher frequency of references to the defendant's ultimate disposition (i.e., what would happen to the defendant following trial). Using disposition-related utterance frequency as the dependent variable, the null model was significant, $\chi(9) = 60.56, p < .001$, but dispositional instructions did not significantly contribute to this difference, $B = 0.28, SE = 0.44, p = .546$. The effect of punishment orientation on disposition-related utterances was non-significant, $B = 0.21, SE = 0.16, p = 0.192$.

## Study 2 Discussion

Whereas Study 1 addressed the role of punishment and insanity attitudes in predicting jurors' initial framing of the case (i.e., their pre-deliberation "story status", Devine, 2012), Study 2 took this one step further by assessing whether attitudes influenced public position-taking. Ten jury deliberation sessions were transcribed and then coded for a-priori content categories by two independent raters. I provided a general breakdown of the content of the deliberation with respect to legal instructions, beliefs about mental disorder, the defendant's ultimate disposition, position-taking, the psychiatrist's trustworthiness, mock juror anecdotes, references to the actual experiment, and utterances unrelated to these categories. Results demonstrated that indicating that the Defence had a strong case or that the Crown had a weak case were rare, while indicating that the Defence had a weak case was common. In sum, questions or opinions about the

actual law (including the burden of proof, reasonable doubt, and the prongs of NCRMD), about appropriate punishment, and about mental disorder in general together accounted for substantial proportions of discussion. There were also a handful of instances in which jurors could not suspend their disbelief, either speculating about the purpose of the experiment or believing that the transcript was fabricated. Finally, mock jurors also relied on personal anecdotes during discussion.

I predicted that permissive punishment orientation would be associated with greater likelihood of advocating for the Crown and lower likelihood of advocating for the Defence. I also expected permissive punishment orientation to predict lower likelihood of advocating for the credibility of the psychiatrist and greater likelihood of referencing the defendant's ultimate disposition. Given that the burden of proof is on the Defence in this case, I anticipated that people higher in permissive punishment orientation would be more likely to reference legal instructions. Results of hierarchical linear modeling did not support the predictions regarding Crown position-taking, nor for references to legal instructions and the defendant's disposition. However, permissive punishment orientation was associated with a decreased likelihood of Defence position-taking and referencing the psychiatrist's credibility. Therefore, it seems that punishment orientation predicts utterances associated with Defence evidence but not Crown evidence.

**Implications**

These findings support previous work on punishment and the insanity defence. In their study of judge's instructions and jury deliberations, Wheatman and Shaffer (2001) found that while including information about the defendant's disposition did not influence individual verdicts, juries who heard these instructions were significantly more

likely to find the defendant NGRI. Thus, consideration of the defendant's disposition can elicit leniency. Results in the current study indicated that dispositional instructions did not significantly contribute to differences in disposition-related utterances. However, the sample size was too small to test for differences in final verdict outcome. In general, results suggest that jurors bring idiosyncratic interpretations of evidence based on pre-existing attitudes. Per Study 1 results, punishment orientation significantly predicted pre-deliberation verdicts; this effect seems to carry over into what jurors actually espouse in a group context.

Previous work on punishment informs the finding that permissive orientation predicted likelihood of public Defence advocacy. Yamamoto and Maeder (2019) argued that lay attitudes do not per se follow the traditional philosophical dichotomy, but rather that people vary in terms of using those principles to justify a desire or aversion to punishment. In short, for some people punishment is rewarding but for others it is itself punishing. Moral reasoning literature demonstrates that people tend to have an instinctual aversion toward harming others, but that this feeling can be overridden by cognitive calculations, as where one can prevent a greater harm (Greene et al., 2001; Valdesolo & DeSteno, 2006). This literature typically features classic utilitarian and deontological dilemmas, in which a moral trade-off is intentionally built to feature an aversive harm to another (e.g., an innocent person). However, Yamamoto and Maeder (2019) argued that punishment decisions are distinct given that by some philosophies harm is not a regrettable necessity but rather desirable and obligatory. The current study suggests that this individual difference of aversion to punishment predicts advocating for leniency. This result also supports Skeem et al. (2004) and Breheney et al.'s (2007) contention that

negative attitudes toward insanity may be a proxy for punishment goals.

It is, therefore, perhaps unsurprising that these attitudes drive leniency rather than harshness. Because the burden is on the Defence to prove insanity, jurors who are punishment-motivated might look the same as those who simply have a high threshold of reasonable doubt. In essence, some jurors might have been able to override an aversion to traditional punishment in favor of following the law, while others might have simply been punishment-motivated. However, those who are willing to make inferences from the evidence that support an insanity narrative may be particularly averse to traditional punishment. Individuals with greater strength of conviction with respect to prohibitive punishment orientation might also have needed to defend their position intensely given the burden of proof.

**Limitations**

Results of Studies 1 and 2 provided support for the role of punishment orientation in informing both jurors' initial survey of the evidence as well as their public advocacy for those positions. Evidently, jurors leveraged pieces of the Crown and Defence's respective narratives in doing so. However, there are a handful of limitations that motivated a closer look at the data, as well as ones that cannot be overcome in this dissertation.

First, the data are underpowered to examine group outcomes. Because the primary interest in this study was descriptive rather than inferential, I did not examine relationships between variables of interest and final verdict outcomes. Second, simplistic descriptions of the number of uninterrupted speaking chains evince a wide variety of talk-time negotiations. Unfortunately, closer examination of the power dynamics therein is

beyond the reach of this study. However, the issue of power and turn-taking is an important direction for future research. For example, Brown (2000) examined gender differences in focus group participation and concluded that failure to account for power dynamics can shift the conclusions of a study. Brown (2000) highlighted that in situations of unstructured turn-taking, men sometimes speak more. Through qualitative analysis, Brown (2000) showed that in an unstructured setting, women tended to offer more limited responses that built off of others. Such gender effects could have accounted for significant patterns in the current study. Previous researchers have shown that women might be more accepting of psychiatric evidence as compared to men (Breheney et al., 2007), and so it is possible that gender effects altered deliberation content. Future research should therefore control for or explicitly examine participant gender in mock juror discussions.

Third, many of the categories from the initial code bank were untenable, and it became necessary to narrow the focus of the initial codebook. Some of this issue is attributable to low frequency of utterance category, as with explicit reference to retributive and utilitarian punishment principles and to certain insanity defence myths. However, this is potentially an indication that these issues do not necessarily play into jurors' justifications of their positions. For instance, although there were some instances of insanity myths, such ideas might be more of an internal process rather than a persuasive tactic. Similarly, in terms of punishment orientation, given previous work showing that punishment philosophy is largely beyond the ken of the average person (Carlsmith et al., 2002; Yamamoto & Maeder, 2019) it makes sense that punishment related utterances would not be as neatly divisible as the coding manual required.

Evidently, the disposition of the defendant is a relevant consideration among mock jurors. However, it is unclear from the quantitative content analysis alone how participants engaged the idea of punishment.

Further, aspects of the Director's cut model were too difficult to quantify without more explicit parameters. For instance, while verdict confidence bears some semblance to story status, (with high confidence in a guilty versus NCRMD verdict representing believers and doubters respectively, and those in the middle representing mullers), criteria for story sampling is difficult to explicate. We might contrast utterance exchanges (i.e., interrupted or terminated turn-taking) containing evidence content to represent isolated statements, with continuations involving evidence (i.e., complete stories). However, the central role of narratives in explanation models of juror decision-making arguably beckons qualitative methods. If the idea is that jurors co-construct meaning in a unique discursive context, then quantitative methods become limiting. It may also be possible to examine the similarities between mock jurors' stories and Defence and Crown opening/closing statements, but this introduces some degree of interpretation into the analysis. Accordingly, I conducted a third exploratory study that relied on inductive methods, to the aim of situating the data in participants' narratives and usage of punishment orientations.

## Study 3 Overview

While the general goal of content analysis is to reduce a large amount of data to a more concise rendering of a phenomenon, as Hsieh and Shannon (2005) summarized, it is also a tool for subjective interpretation. Qualitative content analysis moves further on the interpretive spectrum as compared to Quantitative content analysis (Sandelowski, 2000).

However, Qualitative Description can be considered relatively "low-inference", in contrast to other Qualitative methods such as grounded theory (Sandelowski, 2000, p. 335). The aim of Qualitative Description is simply to present content in everyday terms. In their "how to" guide to qualitative content analysis, Erlingsson and Brysiewicz (2017) depict the process as movement from manifest content to higher levels of abstraction or interpretation. As Sandelowski (2000) argued, researchers conducting quantitative content analysis have less flexibility to understand the variety of unanticipated meanings participants might ascribe to objects. Hence, qualitative description can help to expand on findings from Study 2 while maintaining sufficient philosophical cohesion with the paradigm of this dissertation.

In the summative content analysis approach, keyword and content searchers serve to identify language that is manifestly representative of a construct. Then, moving beyond manifest content, the researcher tries to understand the context in which those terms are mobilized, in attempts to uncover alternative meanings (Hsieh & Shannon, 2005). Hence, this method fills a gap by providing a more nuanced picture of the role that the variables of interest play in persuasion. The purpose of Study 3 was to provide a closer look at the general content identified in Study 2 and to explore usage of language related to moral intuitions (i.e., moral foundations theory, Graham et al., 2009; Haidt, 2009). Taking direction from Hsieh and Shannon (2005), Erlingsson and Brysiewicz (2017), and Sandelowski (2000), I first completed a qualitative description of utterances whose content was about case evidence and the defendant's disposition.

Because some of the content codes from Study 2 appeared to require a greater degree of interpretability, it was more difficult to detect using a manifest content

approach. This resulted in some untrustworthy quality metrics. As a solution, I used the Linguistic Inquiry Word Count (LIWC, Pennebaker, Booth, Boyd, & Francis, 2015) program to flag utterances containing language related to the moral intuitions of authority, fairness, and harm[3]. The moral foundations dictionary, which comprises a collection of words associated with each foundation, has been extensively contextually validated (Graham et al., 2009), and so it was a more reliable gauge of some of the constructs of interest as compared to the initial codes. Graham et al. (2009), with the assistance of a team of researchers, created this dictionary by searching for words theoretically associated with each dimension and then examining use of those words in transcribed religious sermons. The dictionaries for fairness and harm served as a proxy for retributive desires. The authority dictionary served to further probe mock jurors' discussions about the psychiatrist and about the law itself.

## Study 3: Method

I conducted a summative content analysis, in which I coded utterances relating to a-priori content of interest. After using LIWC (Pennebaker et al., 2015) to conduct a keyword search for moral foundations language, I content analyzed all resulting passages, in addition to those flagged as containing discussions about the defendant's disposition and of Crown and Defence evidence.

Following Erlingsson and Brysiewicz (2017), I first read and re-read the utterances from all 10 juries, to get a general overview of the content. I then broke

---

[3] I examined data on all five moral foundations, but the purity and ingroup dimensions only yielded a handful (48 and 55 respectively) of usages, most of which were contextually invalid. Therefore, I only analyzed fairness, harm, and authority. Implications are explored in the Discussion 3 section.

utterances down into meaning units, in a similar fashion to Greene et al. (2008). Meaning units were defined as an utterance by a single juror on a single topic, which had independently substantive meaning. Meaning units were then condensed into smaller representative phrases, still staying as close as possible to the participants' words. I then assigned codes to each utterance, which captured their general essences. As I moved from jury to jury, I re-read previously coded utterances and refined as new ideas emerged. These codes along with exemplar utterances were combined into a code manual, and then conceptually similar codes were placed together. Finally, I assigned categories that captured the relationship between these conceptually similar codes.

**Quality Assurances**

Given that qualitative content analysis moves beyond manifest content to encompass some subjective interpretation, drawing parallels to quantitative methods of quality assurance (e.g., inter-rater reliability) is a complicated endeavour. Debates are abundant with respect to the appropriateness of concepts such as "sameness" or "replicability" in a paradigm in which interpretive differences are celebrated. As Mayring (2014, p. 114) articulated, qualitative content analysis necessitates "modesty" with respect to notions of inter-rater reliability, due this interpretive nature. Mayring (2014) thus recommends that a second researcher serve as quality control by "supervising" and "checking" the first coder's work. Specifically, the analyst must be transparent about inductive coding procedures such that another researcher could prima facie follow the chain of reasoning. This suggestion that the inductive coding process should logically follow, even if the perspective is different than others would choose, echoes across other forms of qualitative analysis. Hsieh and Shannon (2005) similarly highlighted concepts

such as “credibility” and “persistent observation” as metrics of trustworthiness. In short, transparency and clear evidence of logical connection between codes and participant utterances is paramount in qualitative content analysis. Because I stayed relatively toward the descriptive end of the spectrum, codes should reasonably read as present or absent. Notably, researchers also emphasize the importance of presentation in qualitative content analysis, given that “one narrative size does not fit all” (Sandelowski, 1998, p. 376). Per Sandelowski’s (2000) recommended options for presentation, I elected to juxtapose content from guilty and NCRMD voters, in what is termed the Rashomon Effect. Implications for this decision are explicated in the Study 3 Discussion section.

## Study 3: Results

### Evidence

This analysis probed the question of how jurors who voted guilty versus NCRMD mobilized the same evidence differently. Three broad categories of evidence-related utterances emerged: established evidence, simulated evidence, and evidence deficiency. In essence, mock jurors varied with respect to their tolerance of ambiguity in case facts and their willingness to fill in the blanks.

**Established evidence.** A common tactic was for mock jurors to justify their position by simply re-stating established evidence. These utterances often boiled down to whether mock jurors believed the psychiatric testimony; they did not always articulate a reason why they found this evidence compelling. One strategy was to walk through the defendant’s actions, framed in a way that supported the delusion, for example by stating that the defendant took the victim’s wallet because he was under a paranoid delusion. Participants occasionally packaged their utterances as factual or their conclusions self-

evident:

> J10D: Umm.. I guess mostly from the doctor's…, umm…statement and from his point of view, I kinda saw where that was coming from and, umm.., you know the fact that he feels as though people are out to get him, like that says a lot and the fact that he could have acted out, thinking that, you know, whatever he was thinking at that moment that was clearly not…umm, sane .

This factual packaging of the evidence occurred among both NCRMD and guilty voters. For many guilt voters, it was sufficient that the defendant had admitted to the physical act. It was common for participants to list the physical evidence and weigh it against the single witness for the Defence. The following passage contrasts the physical evidence as "proof" while the psychiatric testimony is presented as "opinion". A juror who consistently maintained a guilty stance stated:

> J4A: So it was basically what I think and what all the defense had was a professional's opinion on his assessment with the defendant, and well basically he said that he showed signs of schizophrenia or whatever the mental disorder he has, so compared to the, to the Crown's case, which they had proof of um, he stole money, credit cards, uh they found the m-murder weapon, um, the defendant tried to escape when the police was there through this window, that's a strong case compared to the defendant's case, which the only thing they have was a professional's assessment.

Notably, here the participant translates the psychiatrist's formal diagnosis of paranoid

schizophrenia into “signs of schizophrenia”, an issue that recurred during that particular discussion. Thus, while jurors used grocery-lists of case facts in a tautological manner, there were differences in how that evidence was emphasized. The ways in which jurors simulated missing information provides some insight into the apparent acceptance of certain case facts at face value.

**Simulation.** A handful of narratives consistently emerged across juries with respect to evidence not presented. Overall, there were varying degrees of tolerance for speculation. While one strategy was to fill-in-the-blanks, others countered with the need to stick to the facts.

J3E: Would a butcher’s knife be a kitchen knife or not?

J8E: It’s kind of like a cleaver.

[Laughter]

J2E: I mean we don’t know, did he buy that and never used it until that time or did they regularly use it at the place?

J2E: We don’t know that, can’t just speculate on that.

Some jurors tried to simulate what the defendant was experiencing. A striking difference between those favoring the Defence versus the Crown’s account was how participants simulated his actions: the chaotic nature of delusions or their own likely actions in the scenario. Through this exercise, mock jurors provided a wide range of accounts. For instance, some jurors pointed out that delusions are not necessarily parallel to how events would unfold in reality. Those participants tended to rehash the Defence’s urging not to “fall into the trap” of trying to understand the delusion, because delusions are irrational by nature. Others indicated that his actions would be unlikely in the event of

a real alien encounter.

J6G: But, also, like, if, like, put yourself in this situation. Imagine, like, right now, like, (looks at glass) even in this room through the glass whatever and you have, like, a butcher's knife. If I kill some alien, like I'm not worried trying to wash the butcher knife, like – what the **** – you know, like I'm off it.

While some jurors presented their position as self-evident, others moved through logical chains of reasoning. However, these logical chains required acceptance of an initial premise in line with either the Defence or Crown, for which mock jurors often did not provide a rationale. For example, one case fact that jurors disputed in the trial transcript was the defendant's saying: "Don't let them take me!" when the police arrived. Some participants assumed that the defendant was asking his mother not to let the police take him. Conversely, in the following passage, the participant interprets the defendant's statement as indicative of a delusional state.

J4C: So I also thought he was not criminally responsible and in addition to everything that's kind of been said. There was one specific detail that stood out to me. Even though the prosecutor indicated that the psychiatrist had been looking at him two weeks later, when the police saw him at his house, he seemed to indicate, he said something like "Don't let them take me". So clearly, something is going on at that point of time and he didn't have two weeks to come up with the schizophrenic (inaudible). Unless it was a premeditated murder and he had that idea all along, but again it was

his friend, there's a deeper issue going on there (inaudible).

Here, the mock juror's inference that the defendant would be unlikely to fabricate a mental disorder in that time frame might be persuasive if the statement is interpreted consistent with the Defence narrative. In brief, mock jurors appeared to rely on circular reasoning, in which a specific interpretation of the evidence logically leads to their preferred conclusion. Notably, however, both NCRMD and guilty voters also sometimes relied on tentative, intuitive language in presenting their position (e.g., J6B: "It seems like he knew that it was wrong, um").

**Deficiency.** Given the ambiguity in case facts, a consistent issue for those who initially voted guilty was the lack of evidence with respect to the defendant's history of mental disorder, and a general lack of witnesses for the Defence. In contrast, NCRMD voters countered with the lack of criminal history or documented violent behaviour. Others highlighted the psychiatrist's testimony that the defendant was around the average age of onset for schizophrenia, which would account for the lack of history.

A major point of contention between jurors was the sufficiency of a single psychiatrist (e.g., "one man's opinion") to accurately diagnose someone with schizophrenia. Several jurors remarked on the general scope of evidence (e.g., "There's so much missing!"), indicating that the case had insufficient witnesses, or that they wanted more information. For a handful of jurors, this deficiency interfered with their suspension of disbelief. In one instance, the research assistant was called into the room and asked to provide more evidence. Clearly, jurors varied in the extent to which they felt

the ambiguity necessitated a guilty or NCRMD verdict.

**Disposition**

This analysis probed the question of how mock jurors engaged ideas about the defendant's potential punishment. Five general categories emerged with respect to what would happen to the defendant after the trial: the effectiveness of prisons, the conditions of prisons, the jury's duty in considering punishment, desires for rehabilitation, and desires for incapacitation. Hence, the majority of discussion surrounded utility-based concerns, although there were a handful of references to ideas of just desserts (e.g., "scot-free").

First, several mock jurors indicated that prisons are not rehabilitative, which tended to accompany discussions about the amount of suffering in prisons. One NCRMD voter indicated that a defendant should prefer jail over institutionalization, citing that periods of institutionalization are typically longer than prison sentences in these cases. There were varying beliefs about the conditions of corrections centers.

J5H: Also like, being in jail, is not the worst thing that can happen to you…

J10H: (Inaudible) medication.

J5H: Yea, actually, if they get him help, in jail, then…

J12H: Our prison facilities here are (inaudible). You have to share toenail clippers with forty other guys, it's not a great place to be. But he will have to take his medication.

J7H: But I don't think you can, I don't think you can judge like: "Oh, it's gonna be really hard in jail". Like I can't bring that into, like,

my decision as to whether or not he is. I'm sorry.

The above passage also illustrates a third category, which pertained to the jury's duty in considering the defendant's disposition. Some jurors urged others to "follow the law". Others acknowledged that the defendant's disposition was not legally at issue, but nonetheless discussed the topic. On a jury that received dispositional instructions, the following exchange occurred:

J5I: And, I mean there is also, I mean, the the, um, prison could also be, a, danger to, um, somebody, with, a, mental illness, it can go both ways, but I can't, I'm not sure

J6I: But a murderer is dangerous to a mental facility [laughs]

J4I: (Inaudible)

J5I: I'm not sure, it's appropriate for us to consider sentencing?

[Talking over each other]

J5I: There's a whole, there's a whole thing of, I-I mean, I'm personally, strongly, of the opinion that the prison system needs reform, but like, I still, I mean…

[Talking over each other]

J1I: She did, specifically…

J5I: Yea, so I-I, unfortunately I don't think, we can, we're allowed to, like, consider that, I think we just have to go off of the facts (inaudible) different stage with different deliberations.

Juries that were not dispositionally instructed still espoused the idea that it was

not their duty to consider punishment or treatment.

J1C: The one thing that's guaranteed though if you're sent to the psychiatric facility, he will be treated. Okay? You send him to jail, he's gonna do his time, he's gonna get out, and he'll be back in fruitbat land.

J9C: Yeah, but we're not, we're not debating that.

J2C: Yeah, that's not the issue that (inaudible) to me, it whether not whether*[sic]* or not it's effective, I know it's not effective. It's whether or not this specific individual, if (inaudible) to receive the treatment, they would receive because they have a mental illness or whether they should be put to the prison system because they have all their faculties. Well, I agree with you the prison system does nothing to rehabilitate criminals, but to me that's not the issue here, I'm not trying to change. I wouldn't use this case to try to push an agenda outside the specific case.

This passage also highlights a fifth category pertaining to jurors' desires for incapacitation and protection of society. Relatedly, the timing of the defendant's release was a point of contention. Some juries settled on the idea that the defendant would not be released until safe.

J6E: Do you guys know that hospital, like how does that work?

J8E: I think you stay there until the professionals deem you fit, or get

you under control.

J6E: And then right back into normal clothes?

J8E: I think so. And then, I think, definitely probably back to a psychiatrist.

J6E: [Long, loud sigh and head shake]

J8E: But it's not just like you spend a week and…

J6E: Yeah, it would be a while.

Other jurors maintained a concern that the defendant was a danger to society dependent on their decision. This belief persisted among those who questioned the defendant's motive in avoiding prison. For instance, two jurors who previously expressed concerns that the lawyer coached the defendant had the following exchange:

J2A: So what happens next time?

J4A: Does he kill again? Because he… he's [scare quotes] schizophrenic? Are you guys good with your decision? [Gestures at J3 and J6].

J1A: Yep.

A fifth category pertained to jurors' desires for the defendant's rehabilitation. It was commonplace for mock jurors to either indicate that treatment was needed or to use questions about treatment as a persuasive tactic, which were sometimes successful. One juror attempted to persuade the eventual lone holdout of eight:

J3E: If he were to have paranoid schizophrenia, do you believe that jail is the right thing for him?

J1E: No.

J3E: No, you don't.

J1E: The hospital…

J3E: Right.

J1E: If he was cured somewhere like that.

J3E: So, if I've not misunderstood, if is found guilty he would go to jail, if he was found not criminally responsible and he has a mental illness, he would be going to mental hospital or seeing doctors on a regular basis, is that right?

J1E: I would rather him have a (inaudible).

J3E: But isn't that what Not Criminally Responsible means?

The holdout juror changed positions by the end of the deliberation: "And I do like Not Criminally Responsible. The fact that he would be taken care of."

Notably, there were a handful of utterances indicating an unfavorable attitude toward the language of NCRMD because it implies that the defendant is not punishable (e.g., J4F: "I just don't like how they phrased it…It makes it sound like he's not responsible like he's going to get away like"). There were also notions that NCRMD is a legal loophole (e.g., J4A: "And I also believe sometimes that Defence attorneys will tell that to their clients, umm, ya know?"). These ideas are reminiscent of strict liability and injustice and danger attitudes toward the insanity defence. In brief, mock jurors predominantly espoused concerns about rehabilitation, the conditions in and effectiveness of corrections, incapacitation, and their duty in considering punishment.

**Moral Foundations**

This analysis addressed the question of how different moral foundations were

rejected or accepted in persuading other jurors. The search for words related to the purity and ingroup foundations returned insufficient data for analysis. Therefore, I examined how jurors used language relating to the moral foundations of authority, fairness, and harm in justifying their positions. Of note, LIWC (Pennebaker et al., 2015) separately flags words with positive and negative valences, to give a sense of whether words constituted the foundations as "vices" or "virtues".

**Authority.** LIWC (Pennebaker et al., 2015) yielded 6 utterances containing authority vice language and 64 containing authority virtue language. Both NCRMD and guilty voters leveraged principles of authority acceptance and rejection in their utterances but differed in terms of whose authority should be trusted. Unsurprisingly, NCRMD voters deferred to the authority of the psychiatrist, as demonstrated in the following exchange on a hung jury, in which the eventual lone holdout defended his position:

J1I: I think under the rules we were given we're engaging in a lot of speculation about what's, not in there, again I'm, I'm trusting the psychiatrists' diagnosis, and we're engaging…, In, what, to me, is a lot of tenuous speculation "well maybe the psychiatrist is wrong, maybe there's this, maybe there's that" – I think you guys are – [Talking over each other]

J1I: – insinuating a lot of the things, into, the case.

J5I: I do think it's an obligation of jurors, though, to consider the quality of the evidence, and my opinion was that I mean the psychiatrist doesn't, lack quality, but the quality of the evidence would be vastly improved by even, again, just one corroborating

> witness. And they didn't have that, and that makes me feel like that I, really want to believe them, but I just feel like I can't.

In the above passage, while Juror 1I insists that the authority of the psychiatrist be accepted, Juror 5I rejects the idea in favour of consideration of the Crown evidence. Similarly, one juror objected to reliance on the psychiatrist's authority: "J6F: to me it's weird cuz like we're just kinda like putting a label on this thing and saying it's like ok by like certain authorities uh… things are excusable because of this like black box that we don't know – or that I don't know anything about." Rather than the authority of the psychiatrist, guilty voters tended to defer to the authority of the law. For instance, one juror who maintained a guilty verdict throughout deliberation persisted in the idea that the defendant did not have a diagnosis of schizophrenia: "Under the law, or under the hospital act of Canada, he is not certifiable schizophrenic." It was also common for guilty voters to rely on the phrase "that's the law". This notion tended to accompany discussions about the defendant's punishment.

In terms of alternative usages that did not evince rejection or acceptance of this moral foundation, LIWC (Pennebaker et al., 2015) also flagged utterances related to the role of the Defence lawyer. For instance, jurors debated whether the Defence lawyer would encourage a client to erroneously plead NCRMD: "And then the thing, the case, for th-the lawyer, I mean he can't try to convince someone to like act crazy or lie as the lawyer you know…that gets you disbarred or get into a whole lot of trouble too." Usages also encompassed jurors' attempts to establish their own credibility. Jurors who voted guilty sometimes tried to establish their own authority by indicating they had experience with law or forensics. Like guilty voters, NCRMD voters sometimes established their

own credibility, but more often through experience working with persons with mental disorders. For instance, one juror repeatedly espoused facts about anti-psychotics. Specifically, he suggested that the defendant's anti-psychotic appeared to be working, and that this was evidence of a true mental disorder. NCRMD voters also sometimes downplayed their own experience in deference to others' expertise: "I mean, I'm not the doctor so."

**Fairness.** LIWC (Pennebaker et al., 2015) yielded 39 utterances containing fairness vice language and 11 containing fairness virtue language. One of the most frequently flagged words related to the fairness moral foundation was "bias," which appeared to encompass acceptance of this moral foundation in determining correct action. Mock jurors seemed to converge on the idea that bias was an important consideration but differed on whether it was present. NCRMD voters tended to cite the psychiatrist's professional reputation and neutrality. Guilty voters sometimes communicated the idea that the psychiatrist might be biased, ranging from the potential for anyone to be biased to outright dishonesty. Both NCRMD and guilty voters also called attention to their own potential for bias.

J6D: Well, I guess the bias is like, for me is I hang out with a lot of (inaudible) sick faces (inaudible). For me, like, I deal with a lot of kids where they are treated, they're like (inaudible) they're treated like (inaudible).

J9D: I don't know, I think just focus on that particular question. I mean, you can use all of your prior knowledge or your past experience, of course, but I think that based on what we have and then, that particular question, you

have to decide for yourself. [turns to J10] What about you?

J10D: Yeah, I'm, um, again, I'm totally biased as well, but I'm – you know, based on what we have here, I'm definitely gonna say guilty. Unfortunately…

LIWC (Pennebaker et al., 2015) also captured concessions (e.g., "that's fair"), as well as parroting of legal language. In particular, the term "reasonable" is in the fairness dictionary and was primarily used in discussing the concept of reasonable doubt. Occasionally, jurors indicated that another's position was "reasonable." Finally, "justice" and "justification" tended to appear in discussions of whether it was permissible for one to kill an alien or another person in self-defence. In sum, the keyword search for fairness flagged acceptances of this moral foundation, but not rejections of it.

**Harm.** LIWC (Pennebaker et al., 2015) yielded 195 utterances containing harm vice language and 37 containing harm virtue language. The software flagged every instance of the word "kill", given that it represents a harm against another. Kill-related utterances had roughly three forms. First, both NCRMD and guilty voters occasionally acknowledged the harm done (e.g., "he killed a man"). Second, jurors debated the characteristics of "violent people", for example: "I mean most of them either abuse animals before, or get in fights and assaults before the murder, they don't just…" Third, mock jurors focused on the victim's manner of death. For instance, they attempted to make inferences from the number of stab wounds, which were positioned as indicative of the defendant's emotional state:

J1F: I was – I was gonna say that one of the things I considered was how the victim was killed, in other words was it a… you know, a

shot to the head, was it a single stab wound, but multiple stab wounds to the neck –

J5F: You're either angry or you're scared.

J4F: Well if you're angry you just [inaudible – 9:43]

[All at once – inaudible]

J1F: It's like – stab, stab, stab, I don't know, I – I just felt that –

J3F: Yeah, I hear what you're saying.

Unsurprisingly, guilty and NCRMD voters diverged on who the defendant believed the victim was – an alien or his friend – and by extension, his motive for killing. Several jurors questioned the necessity/excessiveness of killing (i.e., that there were alternative choices). Relatedly, the keyword "protect" emerged in two distinct contexts. First, NCRMD voters tended to favour the narrative that the defendant believed it necessary to kill in order to protect society. Conversely, among guilty voters, the term "protect" arose more so in relation to the need to protect society from further violence. One juror even made explicit reference to the Punishment Orientation Questionnaire:

J4I: [Raises hand] Oh! OK! Right! The reason why I was saying that we should consider that is that, I feel like, if we're unsure, if we are leaning toward one way but we're really unsure, then we should go with the decision that benefits society, like. The questionnaire that we filled out? Online?

J5I: Ohhhh!

J4I: Is it that, if someone's responsible, is it that they get what they deserve? Or is it that, it's what, benefits society most, so I'm just saying if it's to the

> point where we're really indecisive? Then we should go with the decision which, is the most beneficial, or protecting of, people, including, inmates, but, I, just [raises hands] like.

As with the fairness dimension, LIWC (Pennebaker et al., 2015) flagged legal language, specifically several occurrences of "suffering from a mental disorder". The term "suffering" appears in the National Judicial Institute model instructions. Understandably, the law appears to encompass multiple moral foundations.

**Study 3: Discussion**

The intention for Study 3 was to provide a more complete understanding of the general constructs highlighted in Study 2. Following a qualitative description paradigm, I conducted a summative content analysis by examining utterances containing key words relating to moral foundations (Graham et al., 2009) and utterances related to the evidence as well as to the defendant's potential disposition. The data were presented to compare and contrast NCRMD and guilty voters' interpretations of case facts as well as their use of the same values in justifying positions.

Oddo (2013) provided a framework through which to understand how an original text tends to become transformed in a new communicative context. Oddo (2013) identified four broad manifestations of this transformation: deletion, addition, relexicalization, and re-ordering. These modes of alteration to original text parallel jurors' accounts of the evidence, for instance in downplaying or overstating the psychiatric diagnosis. Through a rhetorical discourse analysis, Oddo (2013) also demonstrates two specific speaker strategies. First, in what is termed "enhancement", claims from original content are represented with greater assertion. Second, in the process

of "legitimation", the speaker imbues original content with implied authority and logic. These strategies call to mind mock jurors' varying emphases on established and simulated evidence. Oddo's (2013) framework is also reminiscent of the authority moral foundation, in which jurors emphasized either the authority of the law or the psychiatrist. Unwillingness of some guilty voters to speculate about evidence in some ways runs contrary to the idea of a leniency shift, which deliberation can evoke (Wheatman & Shaffer, 2001). Many participants cited the burden of proof as a reason not to speculate, essentially suggesting that the defendant should not have the benefit of the doubt. It seems that confusion regarding the appropriateness of considering the defendant's disposition might have contributed to this finding. While for some, consideration of punishment might yield unwillingness to vote NCRMD, clearly for others it could elicit prosocial concerns for the defendant's welfare.

**Implications**

As Graham et al. (2009) argued, the form of one's moral persuasions can be more interesting than the simple content. In general, results imply that jurors from different positions relied on similar rhetorical strategies, including logical if/then statements as well as appeals to intuition (e.g., "it just seemed like"). They seem to diverge on the premises presented in supporting those arguments; use of moral foundations related language helped to uncover what values might have motivated those interpretations. In a post outlining how lawyers can incorporate moral foundations language into their closing and opening statements, one litigation consultant remarked: "Without a theme, your case is just information: facts, claims, exhibits, instructions, and witnesses" (Broda-Bahm, 2018). This advice appears to be well-placed, given that moral foundations might feature

in discussion. Moreover, the moral foundations share a connection with the primary factors for evidence consideration in Devine's (2012) Director's Cut Model. For instance, authority might speak to mock jurors' assessment of the credibility of the witnesses and other evidence. Overall, the ways in which mock jurors constructed evidence into varied narratives (or declined to fill the gaps) played a central role in discussion, consistent with story models of juror decision-making (Devine, 2012; Pennington & Hastie, 1986)

In terms of discussion surrounding the defendant's eventual disposition, jurors' utterances encompassed several of the major goals identified by both psychologists and philosophers. However, the majority of these goals pertained to utilitarian concerns of rehabilitation, incapacitation, and deterrence. This finding is in line with research showing that when justifying decisions, people tend to prefer utilitarian reasons to retributive ones (Carlsmith et al., 2002). However, their preoccupation with the poor versus acceptable conditions of prisons shares a kinship with retributive concerns (i.e., the amount of harm the defendant might experience). Passages flagged by the moral foundations dictionary provided some insight into jurors' more retributive driven arguments. Specifically, the harm dimension underscored concerns with the degree of damage done by the defendant (as well as future dangers). Findings seem to suggest that jurors do engage moral intuitions during deliberation. Moreover, appearance of this language provides further evidence of contextual validity for portions of Graham et al.'s (2009) Moral Foundations Dictionary. Indeed, moral foundations related language is embedded into the law itself, which participants often echoed.

**Limitations**

There are a handful of limitations that warrant caution in interpreting findings.

First, it is clear that certain features of the experiment exerted an influence on jury discussions. Some participants appeared to have trouble in suspending disbelief. For instance, mock jurors sometimes questioned whether the transcript was fabricated to be balanced. Further, one participant made explicit reference to the POQ, and others used language that appears in the IDA-R. Because I did not counterbalance the deliberation and questionnaires, it is impossible to say definitively how prior exposure did or did not elicit certain discussion topics. Notably, there was a time delay of three to four days prior to deliberation.

Second, even where qualitative analyses are more descriptive in nature, "immaculate perception" is not possible (Beer as cited in Sandelowski, 2000, p.335). All qualitative analyses feature some amount of interpretive liberties. For instance, I chose to present findings by comparing and contrasting guilty and NCRMD voters, which might have influenced results. As such, the analysis does not focus on jurors who were relatively uncertain, but rather captures those who had stronger moral convictions. However, that interpretation more closely approximates my research question, and so I leave ideas about how uncertainty manifests in terms of swing votes to future researchers.

Similarly, the main question in this thesis pertains to punishment orientation, and so that narrowed focus left a range of other interesting themes unexplored. Specifically, a substantial percentage of utterances related to jurors' beliefs about mental disorder, but I did not qualitatively analyze those ideas. Work by Yamamoto et al. (2017) indicates that mental disorder stigma is a likely source of juror bias against the insanity defence. In future, these data could be examined to understand how beliefs about mental disorder

manifest in a persuasive context.

Further, I was unable to analyze language surrounding two of the five moral foundations (i.e., sanctity and ingroup loyalty) due to low counts and lack of contextual validity. However, the dimensions of fairness, harm, and authority are of particular interest because of the implication that NCRMD defendants are not punishable (fairness), the violence and consequence of the act (i.e., harm), and the centrality of psychiatric testimony (i.e., authority). While the sanctity and loyalty foundations may have application, they appear to have been less central to these particular deliberations.

Finally, while the inevitable interpretability of the data is an issue, by the same token, further interpretation would likely yield a richer picture. Again, my qualitative analysis does not account for the powered dynamic of group discussion (e.g., how jurors actually negotiated turn-taking or monopolized conversation). I also did not delve deeply into the ways that participants might have co-constructed meaning differently across the juries. Discourse analysis or grounded theory might be better suited to tackle such a question.

## General Discussion

This dissertation explored the role of punishment orientation in mock jurors' pre-deliberation decisions and deliberation content in response to a fabricated NCRMD case involving paranoid schizophrenia. Study 1 involved analysis of individual juror decision-making, including a selection procedure (i.e., Challenge for Cause) aimed to identify biased jurors. Study 2 involved a directed content analysis and hierarchical linear models testing the relationship between punishment orientation and deliberation content. Finally, Study 3 more closely examined that content through a combination of key word searches

and qualitative description.

The Introduction opened with a quotation from Judge David Bazelon: "Our collective conscience does not allow punishment where it cannot impose blame". Are jurors likely to agree? Exploring two opposing approaches to moral decision-making, in which peoples' decisions are dominated either by intuitive or rationally calculated processes, my primary interest was what principles and strategies mock jurors relied on in attempting to defend their positions. The rationalist camp might say that punishment should indeed be a calculation based on a defendant's intentionality and of potential future consequences. Conversely, from an intuitionist perspective, one could say that the need for punishment alone imposes blame. Under a retributive framework, for example, justifications for punishment are retrospective; retribution is desired simply by virtue of a harm done. As Tebbit (2002) articulated, retributive oriented people punish because they want to.

Clearly, I cannot presume to answer a century-long debate about the role of intuition versus reasoning in moral decision-making. However, the data can shed light on the relative emphasis certain jurors place on each strategy. In Study 3, I found that mock juror utterances occasionally featured virtually no reasoned or narrative justification of their perspective. Rather, they presented conclusions as self-evident. Mock jurors also evinced acceptance of three moral foundations, in line with Graham et al. (2009). These results lend support to an intuitionist account. On the other hand, jurors also seem to use reasoned arguments surrounding the burden of proof and lack of evidence. For example, several jurors articulated affectively-based reasons for wanting to vote NCRMD (e.g., sympathy), but cited legal instructions as prohibiting reliance on those feelings, which

could be interpreted as a cognitive override of intuitions. Overall, given that punishment orientation shared a significant relationship with juror decisions and utterance frequencies, it is at least clear that jurors do not discount punishment intuitions.

I also presented Devine's (2012) Director's Cut Model to orient the reader to the many potential contributors to juror and jury decisions. Although this was not the central focus of these studies, results of this dissertation do much to recommend this theory. Jurors do appear to rely on pre-existing attitudes in coming to pre-deliberation verdict decisions. They then discuss the scope of evidence and credibility of witnesses, to the aim of articulating a variety of narratives flowing from those ideas. However, it is less clear how one may go about systematically measuring these tendencies using a quantitative approach. In particular, it is difficult to discern how a "puzzler" story status (i.e., inability to make sense out of the case, Devine, 2012) manifests during deliberation. The framework of this dissertation might even suggest that beliefs about insanity are too polarizing to permit puzzling. Rather, mock jurors might instead rely on moral convictions about punishment. More significantly, one could argue that the very notion of narratives in many ways is most consistent with a constructionist lens, because it implies that people engage in meaning-making as the deliberation goes on. It seems that further work is needed in fleshing out operational definitions of various aspects of the Story Sampling Model.

**Implications**

In the Introduction, I articulated three potential responses to the problem of insanity defence bias: change the law, change jurors' minds, or change the jurors. With respect to changing the law, if most jurors regard the insanity defence as itself immoral,

then that might be an indicator that legal reform is necessary. However, results of three studies indicate a wide array of attitudes toward and behaviours surrounding the NCRMD defence. I also noted that the adversarial nature of the system might serve as a safeguard against bias. The variety of opinions on NCRMD and range in tenacity about them supports this idea. However, as the hung juries show, the resiliency of some of these attitudes in response to other jurors' arguments renders this somewhat impractical, albeit thorough.

Further, the Challenge for Cause procedure as conceptualized in this dissertation did not significantly relate to pre-deliberation verdict decisions, although notably endorsement of the insanity defence did. Overall, results imply that the idea of removing biased jurors is overly simplistic. First, the success of the procedure might depend heavily on idiosyncrasies of the case. Moreover, because the case was built to be ambiguous, these data do not distinguish between those who were simply following the law and those who would nullify even in the face of unambiguous evidence.

Thus, it may be left to lawyers and psychiatrists to persuade jurors on the appropriateness of NCRMD. Results of content analyses from Studies 2 and 3 underscore at least three possible strategies. First, regardless of dispositional instructions, mock jurors appear concerned about whether they are permitted or ought to consider the defendant's ultimate disposition. In fact, this dissertation undermines the very notion that jurors are not in the business of deciding punishments. If jurors truly make decisions on the basis of intuitions (Haidt, 2001), then it is difficult to conclude that punishment is irrelevant to the decision. Therefore, perhaps instructions should be amended. Second, lawyers, through opening/closing statements or expert testimony might consider

explicitly depicting either incarceration as rehabilitating or institutionalization as punishing. Third, legal practitioners might consider the role of authority in juror decisions. Study 3 indicated that some NCRMD and guilty voters diverged on deference to the law versus to the psychiatrist. Again, legal instructions explicitly permitting jurors to consider punishment might change decisions. Further, even where there is no opposing expert testimony, a second corroborating expert might be needed. In either case, a strategy might be to feature moral foundations language in opening statements. However, it is worth noting that even in the face of these changes, mock jurors could experience dumbfounding (Haidt, 2001).

**Limitations**

There are a number of general limitations that require consideration. Chief among those limitations is the exploratory nature of the study, which resulted in a small-scale investigation lacking experimental manipulations. Although this restricted focus was intended to produce a rich rather than comprehensive project, several variables of interest were left unexplored. For instance, research underscores mental disorder stigma as a likely source of juror bias, which can vary as a function of mental disorder type (Yamamoto et al., 2017). Likewise, jurors' prototypes about mental disorders seem to influence decision-making (Skeem & Golding, 2001). My studies also cannot account for the intersectionality of experiences with mental disorder (Crenshaw, 1989). It is likely that characteristics of the defendant (e.g., race, gender) change jurors' perceptions. For instance, in a study examining race and mental disorder type in a fabricated NCRMD trial, Maeder, Yamamoto and McLaughlin (under review) found that a Black defendant was more likely to be found guilty in a case involving schizophrenia as compared to one

involving depression. Participant characteristics could similarly influence results. I also did not control for group level influences as previous researchers have done, such as simultaneous versus sequential voting (Davis et al., 1988), leaving the power of normative versus information influences ambiguous. Finally, the models I chose in Study 2 were more simplistic than other possible models. It would be useful in future to treat utterances as nested within jurors, and to add time as a variable. Accounting for the presence or absence of certain topics later in deliberation might give us a sense of whether juries were evidence or verdict driven. Similarly, a time variable could provide information about whether more moralized issues drive deliberation or occur toward the end of discussion after thorough discussion of evidence.

It is also worth acknowledging that conceptualizations of punishment throughout this dissertation are based on Western, individualistic cultural values. Specifically, the POQ (Yamamoto & Maeder, 2019) does not feature a range of other punitive goals. Indigenous Peoples, for instance, might prefer a restorative justice approach to dealing with criminal offences, in which the community works together to prioritize healing (Achtenberg, 2015). Along the same vein, there is a culture in psychology that tends to regard institutionalization as rehabilitative and appropriate. I acknowledge that this perspective has been challenged by critical theorists; some may regard forced institutionalization as problematic and oppressive. At the same time, other researchers may argue ardently that corrections centers are rehabilitative and equipped to handle mental health-related issues. It is necessary to challenge one's ontological and epistemological assumptions in interpreting research findings. However, the data cannot speak to the relative success and harm of the mental health care and criminal justice

systems. Rather, findings speak to the connection between current laws and mock jurors' interactions with them.

Finally, my general approach to coding likely impacted on findings. I elected to code utterances at a small grain-size (i.e., sentences or idea units), but it would also have been possible to analyze longer stretches of text, such as reasoning chains. I also could have segmented utterances into arguments, which would have given a stronger sense of the success of jurors' persuasive tactics. Although these issues are significant, they are also hypothesis generating.

**Future Directions**

While there are limitations to this dissertation, which leave many questions unanswered, there are several promising future directions. This mixed-methods project was intended to be descriptive and exploratory, given the lack of insanity deliberation studies to build from. Thus, there are many opportunities for both qualitative and quantitative follow-up studies.

First, armed with a general sense of the topics that jurors discuss, as well as the attitudes involved, a full-fledged, large-scale quantitative study can be designed. This study would involve a number of experimental manipulations. With respect to features of the trial stimulus, future researchers may wish to manipulate language and/or arguments in lawyers' opening statements. It would also be instructive to manipulate the gender of the defendant as well as type of mental disorder. Further, researchers should consider testing changes to the model disposition instructions, which still feature information about dispositions, but either do or do not discourage jurors from considering that disposition. Additionally, use of confederates would greatly expand this endeavour.

Specifically, confederates could pose certain moral arguments or attempt to dumbfound (Haidt, 2001) mock jurors. Salerno and Peter-Hagene's (2015) computer-mediated deliberation method could be useful for such an investigation. Relatedly, there are several opportunities surrounding participant characteristics. For instance, jury composition could be controlled, such that participants of certain punishment orientations or political orientations comprised majorities or minorities. In brief, insanity deliberations are rich with opportunity for quantitative experimentation.

Second, the qualitative method has much to offer our understanding of NCRMD deliberations. To the aim of philosophical cohesion, I elected to keep qualitative analyses more in the descriptive realm. However, the interpretive description approach (Thorne, Reimer Kirkham, & MacDonald-Emes, 1997) as well as grounded theory (Strauss, 1987) could add greater richness to results. By way of example, Lynch and Haney (2014) conducted a qualitative analysis on the role of emotions in capital jury deliberations. Lynch and Haney (2014) gave an account of how jurors leverage emotional appeals, packaged as rational premises. Given the significant overlap in jurors' attitudes toward the death penalty and toward insanity (Bloechl et al., 2007), such an investigation could illuminate the effectiveness of different persuasion techniques.

There are also a number of mixed-methods follow-up questions that could be examined. In particular, other researchers would benefit from creation of new domain dictionaries. Identification of language that relates to strict liability and injustice and danger might aid in the problem of low base-rates and coder reliability. Existing dictionaries could also facilitate such an undertaking. For example, the LIWC sentiment analysis (Pennebaker et al., 2015), which flags emotion-laden language, could give a

sense of the range of emotions expressed in deliberation. Coders could then conduct contextual validation following Graham et al. (2009).

**Conclusion**

What was in the black box? Results of three studies demonstrate that ideas about appropriate punishment feature in both thought process and persuasion, and that mock jurors' decisions stem partially from moral conceptualizations of insanity rather than the evidence alone. Findings also provide a glimpse into the power of mixed methodology and highlight limitations of the quantitative paradigm in providing a complete understanding of narratives in unique discursive contexts.

**References**

Abelson, R. P., Kinder, D. R., Peters, M. D., & Fiske, S. T. (1982). Affective and semantic components in political person perception. *Journal of Personality and Social Psychology, 42*(4), 619-630.

Adams, K. & Ferrandino, J. (2008). Managing mentally ill inmates in prisons. *Criminal Justice and Behaviour, 35*(8)*,* 913-927. doi: 10.1177/0093854808318624

Allport, G. W. (1979). *The Nature of Prejudice*. Cambridge, MA: Addison Wesley. (Original work published 1954).

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Angermeyer, M.C, & Dietrich, S. (2006). Public beliefs about and attitudes towards people with mental illness: A review of population studies. *Acta Psychiatrica Scandinavica, 113*(3), 163-179. doi: 10.1111/j.1600-0447.2005.00699.x

Asch, S. E. (1951). Effects of group pressure on the modification and distortion of judgments. In Guerwitz (ed.), *Groups, Leadership and Men* (pp. 177-190). Pittsburgh, PA: Carnegie.

Banaji, M.R., & Heiphetz, L. (2010). Attitudes. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of Social Psychology Volume 1, 5th Ed* (pp. 353-393). John Wiley & Sons Inc, Hoboken, NJ.

Bartels, D., & Pizzaro, D. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121*, 154-161.

Bentham, J. (2006). An introduction to the principles of morals and legislation. In S. Cahn, & P. Markie (Eds.), *Ethics – History, Theory, and Contemporary Issues* 3rd

Ed (pp. 309- 316). Oxford University Press. (Original work published 1789).

Bloech, A.L., Vitacco, M.J., Neumann, C.S., & Erickson, S.E. (2007). An empirical investigation of insanity defense attitudes: Exploring factors related to bias. *International Journal of Law and Psychiatry, 30*, 153-161. doi:http://dx.doi.org/10.1016/j.ijlp.2006.03.007

Bluhm, R. (2014). No need for alarm: A critical analysis of Greene's dual-process theory of moral decision-making. *Neuroethics, 7*, 299-316.

Breheney, C., Groscup, J., & Galietta, M. (2007). Gender matters in the insanity defense. *Law and Psychology Review*, *31*, 93-124.

Brickman, P., Rabinowitz, V. C., Karuza, J., Jr., Coates, D., Cohn, E., & Kidder, L. (1982). Models of helping and coping. *American Psychologist, 37*, 368-384.

Butler, B. M., & Moran, G. (2002). The role of death qualification in venirepersons' evaluations of aggravating and mitigating circumstances in capital trials. *Law and Human Behavior, 26*(2), 175-184. doi:http://dx.doi.org/10.1023/A:1014640025871

Canadian Charter of Rights and Freedoms, PART I OF THE CONSTITUTION ACT, 1982.

Canadian Criminal Code, R. S. 1985, c. C-46, s. 638; R. S., 1985, c. 27 (1st Supp.), s. 132, c. 31 (4th Supp.), s. 96; 1997, c. 18, s. 74; 1998, c. 9, s. 6.

Charette, Y., Crocker, A. G., Seto, M. C., Salem, L., Nicholls, T. L., & Caulet, M. (2015). The National Trajectory Project of individuals found not criminally responsible on account of mental disorder in Canada. Part 4: Criminal recidivism. *The Canadian Journal of Psychiatry / La Revue Canadienne De Psychiatrie, 60*(3),

127-134.

*Clark v. Arizona,* 548 U.S. 735 (2006)

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement 20*, 37-46.

Cohen, T. (2013) Emotional Stephen Harper tears up over controversial bill to keep mentally ill murderers in prison longer. *National Post*. Retrieved from http://news.nationalpost.com/news/canada/emotional-stephen-harper-tears-up-over-controversial-bill-to-keep-mentally-ill-murderers-in-prison-longer

*Colgrove v. Battin,* 413 U.S. 149 (1973)

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology, 104*(2), 216-235.

Criminal Code of Canada, RSC 1985, c C-46 s. 649.

Criminal Code of Canada, R.S., c.C-34, S.16; 1991, c.43, s.2

Crocker, A. G., Nicholls, T. L., Seto, M. C., Charette, Y., Côté, G., & Caulet, M. (2015). The National Trajectory Project of individuals found not criminally responsible on account of mental disorder in Canada. Part 2: The people behind the label. *The Canadian Journal of Psychiatry / La Revue Canadienne De Psychiatrie, 60*(3), 106-116.

Crocker, A. G., Nicholls, T. L., Seto, M. C., Côté, G., Charette, Y., & Caulet, M. (2015). The National Trajectory Project of individuals found not criminally responsible on account of mental disorder in Canada. Part 1: Context and methods. *The Canadian Journal of Psychiatry / La Revue Canadienne De Psychiatrie, 60*(3),

98-105.

Davis, J. H. (1973). Group decision and social interaction: A theory of social decision schemes. *Psychological Review, 80*, 97-125.

Davis, J. H., Stasson, M. F., Ono, K., & Zimmerman, S. (1988). Effects of straw polls on group decision making: Sequential voting pattern, timing, and local majorities. *Journal of Personality and Social Psychology, 55*(6), 918-926.

Day, E.N., Edgren, K., & Eshelman, A. (2007). Measuring stigma toward mental illness: Development and application of the mental illness stigma scale. *Journal of Applied Social Psychology, 37*(10), 2191-2219. doi: 10.1111/j.1559-1816.2007.00255.x

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

Devine, D. J. (2012). *Jury Decision-Making: The State of the Science.* New York and London: New York University Press.

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), 109-134. doi:http://dx.doi.org/10.1037/law0000006

Devine, P.G., & Elliot, A.J. (1995). Are racial stereotypes really fading? *The Princeton Trilogy revisited. Personality and Social Psychology Bulletin, 21,* 1139-1150

Devine, D. J., Olafson, K. M., Jarvis, L., Bott, J. L., Clayton L.D., & Wolfe, J.T. (2004). Explaining jury verdicts: Is leniency bias for real? *Journal of Applied Social*

*Psychology, 34*, 2069-2098.

Dwyer, S. (2009). Moral dumbfounding and the linguistic analogy: Methodological implications for the study of moral judgment. *Mind and Language, 24*(3), 274-296.

Ellsworth, P. C. (1989). Are twelve heads better than one? *Law and Contemporary Problems, 52*, 207-224.

Faulstich, M. E. (1984). Effects upon social perceptions of the insanity plea. *Psychological Reports, 55*, 183–187. http://dx.doi.org/10.2466/pr0.1984.55.1.183

Feigenson, N., & Park, J. (2006). Emotions and attributions of legal responsibility and blame: A research review. *Law and human Behavior, 30*, 143-161.

Finkel, N.J. (2000). But it's not fair! Commonsense notions of unfairness. *Psychology, Public Policy, and Law, 6*(4), 898-952. doi: 10.1037///1076-8971.6.4.898.

Finkel, N., & Handel, S. (1988). Jurors and insanity: Do test instructions instruct? *Forensic Reports, 1*, 65-79.

Fiske, S. (1998). Stereotyping, prejudice, and discrimination. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp.357-411). New York, NY: McGraw-Hill.

Fitzgerald, R., & Ellsworth, P. C. (1984). Due process vs. crime control: Death qualification and jury attitudes. *Law and Human Behavior, 8*(1-2), 31-51.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 5*, 5-15.

Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? Operationalising data

saturation for theory-based interview studies. *Psychology and Health, 25*(10), 1229-1245.

Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. *The Qualitative Report, 20*(9), 1408-1416.

Gaertner, S. L., & Dovidio, J. F. (2005). Understand and addressing contemporary racism: From aversive racism to the common in-group identity model. *Journal of Social Issues, 61*(3), 615-639. doi:10.1111/j.1540-4560.2005.0024.x

Goleman, D. (1995). Anatomy of an Emotional Hijacking. In *Emotional Intelligence.* New York, New York: Bantam Dell.

Government of Canada. The Human Face of Mental Health and Mental Illness in Canada. 2006. © Minister of Public Works and Government Services Canada, 2006 Cat. No. HP5-19/2006E ISBN 0-662-43887-6

Greene, J. (2009). Dual process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology, 45*, 581-854.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgement. *Psychological Review, 108*, 814–834

Haney, C. (1997). Commonsense justice and capital punishment: Problematizing the "Will of the People". *Psychology, Public Policy, and Law, 3*(2/3), 303-337.

Hans, V. P. (1986). An analysis of public attitudes toward the insanity defense. *Criminology, 4*(2), 393-415.

Hsieh, H., & Shannon, S. E. (2005). Three approaches to qualitative content analysis.

*Qualitative Health Research, 15*(9), 1277-1288.

Hume, D. (2000). *A Treatise of Human Nature*. In D. F. Norton, & M. J. Norton (Eds.), *Oxford Philosophical Texts.* New York: Oxford University Press. (Original work published 1738).

Jackson, L. A., Hodge, C. N., Gerard, D. A., Ingram, J. M., Ervin, K. S., & Sheppard, L. A. (1996). Cognition, affect, and behavior in the prediction of group attitudes. *Personality and Social Psychology Bulletin, 22*(3), 306-316.

Juries Act, R.S.O. 1990, c. J.3

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263-291. doi: 10.2307/1914185

Kalven, H., & Zeisel, H. (1966). *The American Jury*. Chicago: University of Chicago Press.

Kant, I. (2006). Groundwork of the metaphysics of morals. In S. Cahn, & P. Markie (Eds.), *Ethics – History, Theory, and Contemporary Issues* 3rd Ed (pp. 270- 308). Oxford University Press. (Original work published 1785).

Kerr, N. L., & MacCoun, R. J. (1985). The effects of jury size and polling method on the process and product of jury deliberation. *Journal of Personality and Social Psychology, 48*, 349-363.

Knobe , J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63*, 190-193.

Kohlberg, L. (1981). *The Philosophy of Moral Development*. San Francisco, CA: Harper & Row

Liebeck v. McDonald's Restaurants, P.T.S., Inc., No. D-202 CV-93-02419, 1995 WL

360309 (Bernalillo County, N.M. Dist. Ct. August 18, 1994).

Lorge, I., & Solomon, H. (1955). Two models of group behavior in the solution of eureka-type problems. *Psychometrika, 20*, 139-148.

Maeder, E.M., & Laub, C.E. (2012). Changing minds: The effect of course and teaching approach on attitudes toward the legal system. C*riminal Justice Studies, 25*, 17-31. doi:http://dx.doi.org/10.1080/1478601X.2012.657900

Maeder, E. M., Yamamoto, S., & Fenwick, K. L. (2015). Educating Canadian jurors about the not criminally responsible on account of mental disorder defence. *Canadian Journal of Behavioural Science, 47*(3), 226-235.

Mill, J. S. (2008). *On Liberty*. A. S. Kahan (Ed.). Boston: Bedford. (Original work published 1859).

Millar, M. G., & Tesser, A. (1986). Effects of affective and cognitive focus on the attitude–behavior relation. *Journal of Personality and Social Psychology, 51*(2), 270-276. doi:http://dx.doi.org/10.1037/0022-3514.51.2.270

Ministry of the Attorney General. (2015). The jury roll Process. Retrieved from: https://www.attorneygeneral.jus.gov.on.ca/english/courts/jury/geninfo.php

Mouter, N., & Vonk Noordegraaf, D. (October, 2012). Intercoder reliability for qualitative research: You win some, but do you lose some as well? *TRAIL Research School.*

Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin, 83*(4), 602-627.

National Judicial Institute. (2014). Model jury instructions. Retrieved from https://www.nji-inm.ca/index.cfm/publications/model-jury-

instructions/?langSwitch=en

Neises, M.L., & Dillehay, R.C. (1987). Death qualification and conviction proneness: Witt and Witherspoon compared. *Behavioral Sciences and the Law, 5*, 479-494.

Olley, M. C., Nicholls, T. L. & Brink, J. (2009). Mentally ill individuals in limbo: Obstacles and opportunities for providing psychiatric services to corrections inmates with mental illness. *Behavioral Sciences & Law, 27*(5), 811–831. doi: 10.1002/bsl.899

Paulhus, D. (1991). Measurement and control of response bias. In J. P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes* (pp. 17-59). San Diego, CA: Academic Press. Inc.

Paulhus, D. L., & Vazire, S. (1991). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 224-239). New York: Guilford.

Penney, S. R., Morgan, A., & Simpson, A. I. F. (2013). Motivational influences in persons found not criminally responsible on account of mental disorder: A review of legislation and research. *Behavioral Sciences & the Law, 31*(4), 494-505.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology: Learning, Memory, and Cognition, 51,* 521-533. doi: 10.1037/0022-3514.51.2.242

Pendrod, S., & Heuer, L. (1998). Improving group performance: The case of the jury. In R. S. Tindale et al., (Eds.), *Theory and Research on Small Groups* (pp. 127-152).

New York: Plenum Press.

Pendrod, S., & Heuer, L. (1997). Tweaking commonsense: Assessing aids to jury decision-making. *Psychology, Public Policy, and Law, 3*, 259-284.

Perry, S. P., Murphy, M. C., & Dovidio, J. F. (2015). Modern prejudice: Subtle, but unconscious? the role of bias awareness in whites' perceptions of personal and others' biases. *Journal of Experimental Social Psychology, 61*, 64-78.

Pettigrew, T. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin, 5*(4), 461-476.

Pruitt, D. G. (1971). Choice shifts in groups discussion: An introductory review. *Journal of Personality and Social Psychology, 20*, 339-360

*Queen v. M'Naghten*. 8 Eng. Rep. 718 (1843).

*R v. Morgentaler* (1988), 1 SCR 30, 44 DLR (4th) 385

*R. v. Pickton,* 2010 SCC 32, [2010] 2 S.C.R. 198

*R. v. Sherratt*, (1991) 1 S.C.R. 509

R. v. Swain, (1991) 1 S.C.R. 933

*Regina v. Parks* (1993) 15 O.R. (3d) 324 (C.A.).

*Regina v. Williams* (1998) 1 S.C.R. 1128, S.C.J. 49.

Roesch, R., Ogloff, J.R.P, Hart, S.H., Dempster, R.J., Zapf, P.A., & Whittemore, K.E. (1997). The impact of Canadian criminal code changes on remands and assessments of fitness to stand trial and criminal responsibility in British Columbia. *Canadian Journal of Psychiatry, 42*(5), 509-514.

Rosenberg, M. J. and Hovland, C. I. (1960). Cognitive, affective and behavioral

components of attitudes. In Rosenberg, M. J. and Hovland, C. I. (Eds.), *Attitude Organization and Change: An Analysis of Consistency Among Attitude Components*. New Haven: Yale University Press.

Ross, L. (1977). The intuitive psychologist and his shortcomings, In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (pp.173-220). Orlando, FL: Academic Press.

Saks, M. J., & Marti, M. W. (1997). A meta-analysis of the effects of jury size. *Law and human Behavior, 21*, 451-466.

Salerno, J. M., & Diamond, S. S. (2010). The promise of a cognitive perspective on jury deliberation. *Psychonomic Bulletin & Review, 17*(2), 174-179.

Salerno, J. M., & Peter-Hagene, L. (2015). One angry woman: Anger expression increases influence for men, but decreases influence for women, during group deliberation. *Law and Human Behavior, 39*(6), 581-592.

Saltzstein, H. D., & Kasachkoff, T. (2004). Haidt's moral intuitionist theory: A psychological and philosophical critique. *Review of General Psychology, 8,* 273–282.

Sandys, M., & Dillehay, R. C. (1995). First-ballot votes, predeliberation dispositions, and final verdicts in jury trials. *Law and Human Behavior, 19*, 175–195.

Schedler, G. (1980). Can retributivists support legal punishment? *The Monist,* 331-334.

Schuller, R. A., Erentzen, C., Vo, A., & Li, D. (2015). Challenge for cause:

Bias

screening procedures and their application in a Canadian courtroom. *Psychology, Public Policy, and Law, 21*(4), 407-419.

Schuller, R. A., Kazoleas, V., & Kawakami, K. (2009). The impact of prejudice screening procedures on racial bias in the courtroom. *Law and Human Behavior, 33*, 320-328. DOI 10.1007/s10979-008-9153-9

*Shannon v. United States*, 114 S. Ct. 2419 (1994).

Silver, E., Cirincione, C., & Steadman, H.J. (1994). Demythologizing inaccurate perceptions of the insanity defense. *Law and Human Behavior, 18*(1), 63-70. doi: 10.1007/BF01499144

Skeem, J. L., Louden, J. E., & Evans, J. (2004). Venireperson`s attitudes toward the insanity defense: Developing, refining, and validating a scale. *Law and Human Behavior, 28*(6), 623-648. doi:http://dx.doi.org/10.1007/s10979-004-0487-7

Slinger, E and Roesch, R. (2010). Problem-solving courts in Canada: A review and a call for empirically-based evaluation methods. *International Journal of Law and Psychiatry, 33*, 258-264

Sloat, L. M., & Frierson, R. L. (2005). Juror knowledge and attitudes regarding mental illness verdicts. *The Journal of the American Academy of Psychiatry and the Law, 33,* 208–213.

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), 597-612. doi:10.1037/0022-3514.90.4.597

Tebbit, M. (2005). Theories of punishment. *Philosophy of Law: An Introduction* (pp.

192-212). New York, NL: Routledge.

Tetlock, P. E. (1983). Accountability and complexity of thought. *Journal of Personality and Social Psychology, 45*(1), 74–83.

The Not Criminally Responsible Reform Act. (2013). 1st Reading November 25, 2013, 41st Parliament, 2nd session.

*United States v. Neavill*, 868, F.2d 1000 (1989).

Van Bavel, J. J., & Cunningham, W. A. (2010). A social neuroscience approach to self and social categorisation: A new look at an old issue. *European Review of Social Psychology, 21*(1), 237-284.

Vidmar, N., & Miller, D. (1980). Socialpsychological processes underlying attitudes toward legal punishment. *Law & Society Review, 14*(3), 565-602.

Vitacco, M.J, Malesky, L.A, Erickson, S.K., Leslie, W., Croysdale, A., & Bloechl, A. (2009). Measuring attitudes toward the Insanity defense in venirepersons: Refining the IDA-R in the evaluation of juror bias. *International Journal of Forensic Mental Health, 8,* 62-70. doi: 10.1080/14999010903014754

Weiner, B. (1986). *An Attributional Theory of Motivation and Emotion*. New York: Springer-Verlag.

Weiner, B. (2006). *Social Motivation, Justice, and the Moral Emotions* (pp. 125-159). Mahwah: Lawrence Erlbaum Associates.

Weiner, B., Graham, S., & Reyna, C. (1997). An attributional examination of retributive versus utilitarian philosophies of punishment. *Social Justice Research, 10*, 431–452. http://dx.doi.org/10.1007/BF02683293

Weiner, B., Perry, R. P., & Magnusson, J. (1988). An attributional analysis of reactions to stigmas. *Journal of Personality and Social Psychology, 55*(5), 738-748.

Wheatman, S. R., & Shaffer, D. R. (2001). On finding for defendants who plead insanity: The crucial impact of dispositional instructions and opportunity to deliberate. *Law and Human Behavior, 25*, 167–183. http://dx.doi.org/10.1023/A:1005645414992

Whittemore, K. E. & Ogloff, J. R. (1995). Factors that influence jury decision-making. *Law and Human Behavior, 19*, 283–303.

Wiener, R. L., Bornstein, B. H., & Voss, A. (2006). Emotion and the law: A framework for inquiry. *Law and Human Behavior, 30*(2), 231-248.

*Williams v. Florida*, 90 S. Ct. 1893 (1970).

*Winko v. British Columbia* (Forensic Psychiatric Institute), 2 S.C.R. 625 (1999).

Yamamoto, S., & Maeder, E. M. (2019). Creating the Punishment Orientation Questionnaire: An item response theory approach. *Personality and Social Psychology Bulletin. https://doi.org/10.1177/0146167218818485*

Yamamoto, S., Maeder, E. M., & Fenwick, K. L. (2017). Criminal responsibility in Canada: Mental disorder stigma education and the insanity defense. *International Journal of Forensic Mental Health, 16*(4), 313-335. DOI:10.1080/14999013.2017.1391357

Young, R. L. (2004). Guilty until proven innocent: Conviction orientation, racial attitudes, and support for capital punishment. *Deviant Behavior, 25*(2), 151-167 doi:10.1080/01639620490266916

# Appendices

## Appendix A Phase 1 Materials

### Recruitment Notice

We are recruiting Canadian jury-eligible participants for a study "Jury Deliberation and the Not Criminally Responsible on Account of Mental Disorder Defence". Participants must be Canadian citizens, at least 18 years of age, who have not been convicted of an indictable offence without receiving an official pardon. Participants will first be asked to fill out a brief online questionnaire (5 minutes). Participants will be asked to later come to Carleton University and read a transcript from a trial including an allegation of murder, then take part in a jury deliberation with up to 14 other participants. This study is being conducted by Dr. Evelyn Maeder from the Institute of Criminology and Criminal Justice at Carleton University, and PhD student Susan Yamamoto from the Department of Psychology at Carleton University.

The study will take approximately 75 minutes and participants will be awarded $40 in person. Participants must be available to travel to Carleton University on a Saturday. This study has been cleared by the Carleton University Research Ethics Board – B (#xx-xxx).

Interested participants should click the link below for eligibility and further information.

Click here to begin.

## Informed Consent Phase 1

The purpose of an informed consent is to make sure that you understand the purpose of the study and your involvement as a participant. The informed consent must include enough information regarding the study for you to be able to make a well-informed decision regarding whether or not you would like to partake in this study.

**Title**: Jury Deliberation and the Not Criminally Responsible on Account of Mental Disorder Defence

**Research Personnel:** This study is being conducted by Dr. Evelyn Maeder from the Institute of Criminology and Criminal Justice (evelyn.maeder@carleton.ca Tel. 613-520-2600 ext. 2421) at Carleton University in Ottawa, Ontario, Canada, and PhD student Susan Yamamoto (susan.yamamoto@carleton.ca) from the Department of Psychology.

**Concerns:** Should you have any ethical concerns regarding this study, please contact Dr. Shelley Brown (Chair, Carleton University Research Ethics Board – B, Shelley_brown@carleton.ca, 613-520-2600 ext. 1505). For other concerns, please contact the Research Compliance Office at ethics@carleton.ca.

**Purpose:** This is a study to evaluate opinions about the Not Criminally Responsible on Account of Mental Disorder Defence and about punishment in general.

**Task Requirements:**

Participants must be Canadian citizens, at least 18 years of age, who have not been convicted of an indictable offence without receiving an official pardon. Please note that participants found to be ineligible for the study will not receive compensation.

**Phase 1.** *This is part one of a two-part study.* If you agree to participate in part one, then following this consent form you will be asked to answer some demographics questions to ensure eligibility, as well as two questionnaires (40 items, approximately 5 minutes). There is no compensation for answering these two questionnaires. To receive compensation you must take part in both part 1 and part 2.

**Phase 2.** The second phase of the study will take place at Carleton University. Participants will be asked to read a short trial transcript and to have a group discussion (no longer than 45 minutes) with up to 14 other participants, which will be video/audio recorded and observed through a two-way mirror. When you arrive at the lab we will need to obtain your consent to be video/audio recorded. Compensation for participating in both phases of the study is $40 in cash, which will be provided in person immediately following participation in Phase 2 at Carleton University. Full participation in Phase 2 of the study will take up to approximately 75 minutes.

**Potential Risk/Discomfort**: For part 1 you will be asked to answer questions regarding your attitudes about the insanity defence and about punishment in general, and you may become uncomfortable with the nature of the questions. You may refrain from answering any questions on the questionnaire if you are uncomfortable or otherwise do not want to and you will still be eligible to participate in Phase 2 of the study.

**Right to Withdraw:** In addition, at any time, you may discontinue your involvement in this study without penalty. However, if you withdraw prior to phase 2 of the study, we cannot provide you with compensation.

**Anonymity/Confidentiality:** The data collected are strictly confidential. Your name is not associated with the responses you provide. We collect data through the software Qualtrics, which uses servers with multiple layers of security to protect the privacy of the data (e.g., encrypted websites and pass-word protected storage). Please note that Qualtrics is hosted by a server located in the USA. The United States Patriot Act permits U.S. law enforcement officials, for the purpose of an anti-terrorism investigation, to seek a court order that allows access to the personal records of any person without that person's knowledge. In view of this we cannot absolutely guarantee the full confidentiality and anonymity of your data. With your consent to participate in this study you acknowledge this. These data will only be used for research at Carleton University. After completing Phase 1 you will be asked to provide your email address in a separate survey. However, because we will need to correspond by email and assign you a juror code, it is possible that your email address will be associated with your responses until up to 2 weeks after Phase 2 of the study. This information will be kept strictly confidential and we will only use your email address for the purposes of providing information about the study.

This study has been cleared by the Carleton University Carleton University Research Ethics Board – B (#xx-xxx).

- *I have read the above form and understand the conditions of my participation. My participation in this study is voluntary, and I understand that if at any time I wish to leave the experiment, I may do so without having to give an explanation and with no penalty whatsoever. Furthermore, I am also aware that the data gathered in this study are confidential with respect to my personal identity.* ***By checking this box, I'm indicating that <u>I agree to participate in this study.</u>***

- *I have read the above form and understand the conditions of my participation. My participation in this study is voluntary, and I understand that if at any time I wish to leave the experiment, I may do so without having to give an explanation and with no penalty whatsoever. Furthermore, I am also aware that the data gathered in this study are confidential with respect to my personal identity.* ***By checking this box, I'm indicating that <u>I disagree to participate in this study.</u>***

**Continue to survey =>**

**Eligibility Questionnaire**

Thank you for your interest in our study! Before we can schedule you for a session, we just need to make sure that you are eligible to participate. Please answer the following questions:

- Are you a Canadian citizen?
  Yes___ No___
- Are you 18 years of age or older?
  Yes___ No___
- Have you ever been convicted of an indictable offence? If you have been convicted but you received a formal pardon, please select 'No'. Indictable offenses are different from more minor "summary" offences (which carry sentences of fewer than 6 months and fines of less than $2000). Examples include theft, treason, murder, piracy, and robbery.
  Yes___ No___

**Next =>**

---

**Ineligible Message**

I'm sorry; it looks as though you are not eligible to participate in this study. We are looking for participants who could get called for jury duty in Canada (i.e., Canadian citizens at least 18 years of age, who have not been convicted of an indictable offence).

Thank you very much for your time and interest. If you have any questions please don't hesitate to contact us: carletonjury@gmail.com.

**Eligible Message**

Thank you for your response. It looks as though you are eligible to participate in this study. If you would like to participate, you will be asked to travel to Carleton University, at which point you will read a brief trial transcript (about 15 minutes). You will then be asked to discuss the case and reach a unanimous verdict with up to 14 other mock jurors. Full participation in this study will take approximately 75 minutes. The deliberation session will be audio and video recorded. These data will be kept strictly confidential (stored offline in a locked lab facility). We will need to obtain your consent for this when you arrive at the lab.

Before you come to the lab, we will need you to fill out 2 brief questionnaires. To receive the $40 cash compensation, we will need you to fill out these questionnaires *and* travel to Carleton for the study; there is no compensation for the questionnaires alone. After you fill out these questionnaires, you will receive a confirmation email for the deliberation study.

We have the following sessions available. Please select one:

**Proceed to questionnaires =>**

**Insanity Defence Attitudes-Revised Scale (IDA-R)**

In this section, you will find statements that express commonly held opinions about the not criminally responsible defence. We would like to know how much you agree or disagree with each of these statements. You may interpret the seven points on this scale as follows:

**1 2 3 4 5 6 7**

**Strongly disagree Neutral Strongly agree**

1. I believe that people should be held responsible for their actions no matter what their mental condition.
2. For the right price, psychiatrists will probably manufacture a "mental illness" for any criminal to convince the jury that he is not criminally responsible
3. I believe that we should punish a person for a criminal act only if he understood the act as evil and then freely chose to do it.
4. I believe that all human beings know what they are doing and have the power to control themselves.
5. The not criminally responsible defence threatens public safety by telling criminals that they can get away with a crime if they come up with a good story about why they did it.
6. I believe that mental illness can impair people's ability to make logical choices and control themselves.
7. A defendant's degree of mental illness is irrelevant: if he commits the crime, then he should do the time.
8. The not criminally responsible defence returns disturbed, dangerous people to the streets.
9. Mentally ill defendants who plead not criminally responsible have failed to exert enough willpower to behave properly like the rest of us. So, they should be punished for their crimes like everyone else.
10. As a last resort, defence attorneys will encourage their clients to act strangely and lie through their teeth to appear "mentally ill."
11. Perfectly sane killers can get away with their crimes by hiring high-priced lawyers and experts who misuse the not criminally responsible defence.
12. The not criminally responsible plea is a loophole in the law that allows too many guilty people to escape punishment
13. We should punish people who commit criminal acts, regardless of their degree of mental disturbance.
14. It is wrong to punish people who commit crime for crazy reasons while gripped by uncontrollable hallucinations or delusions.
15. Most defendants who use the not criminally responsible defence are truly mentally ill, not fakers.
16. Some people with severe mental illness are out of touch with reality and do not understand that their acts are wrong. These people cannot be blamed and do not deserve

to be punished.
17. Many of the crazy criminals that psychiatrists see fit to return to the streets go on to kill again.
18. With slick attorneys and a sad story, any criminal can use the not criminally responsible defence to finagle his way to freedom.
19. It is wrong to punish someone for an act they commit because of any uncontrollable illness, whether it be epilepsy or mental illness.

**The Punishment Orientation Questionnaire (POQ)**

In this section you will find statements about punishment in the criminal justice system. We would like to know how much you agree or disagree with each of these statements. You may interpret the five points on this scale as follows:

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| **Strongly disagree** | | **Neutral** | | **Strongly agree** |

**Prohibitive Utilitarian Scale**

1. Punishment should be about looking forward to improve society, not backward to address the criminal's misdeeds.
2. When considering an appropriate punishment, the potential benefit to the public is more important than the need to avenge the particular injustice.
3. Punishment is more about addressing society's needs than serving out justice to a single individual.
4. We should try to focus on how punishment can help the community instead of fixating on one person's wrongdoings.
5. When punishing a person, it is better to take a step back to think about how that punishment will affect the community.

**Prohibitive Retributive Scale**

1. It is better to let 10 guilty criminals go free than to punish 1 innocent person.
2. It is more important to keep innocent people free from punishment than it is to ensure that all guilty persons are punished for their crimes.
3. A punishment system should prioritize bringing offenders to justice, even at the risk of punishing the wrong person. (Reverse code)
4. Catching more guilty people isn't worth the expense of false convictions.

**Permissive Utilitarian Scale**

1. An overly harsh punishment may be necessary to prevent others from committing the same crime.
2. If a crime has a low detection rate (i.e. it is difficult to catch criminals who commit this particular crime), we should punish those who are caught harshly to prevent others from thinking they can get away with it.
3. Crimes that receive a great deal of publicity should be punished severely, even if the crime was not severe, so that society knows there is a strong response.
4. We should err on the side of stricter punishments if it will result in greater public safety.

**Permissive Retributive Scale**

1. Even if society would not benefit at all from punishing a guilty person, he should still be punished because he deserves it.
2. Criminals are bad people and get what is coming to them.
3. Punishment is necessary because it restores the balance of justice.
4. It is more important to punish a guilty person because he deserves it than it is to punish

him to benefit society.

## Challenge for Cause

Please respond to the following questions and note that your answers will not affect your eligibility for participating in the remainder of the study.

1. Do you believe that the insanity defence should be abolished?
    Yes.
    No.

2. Would your ability to judge the evidence in this case without bias, prejudice, or partiality be affected by the fact that the accused is pleading Not Criminally Responsible on Account of Mental Disorder?
    Yes.
    No.

## Appendix B Phase 2 Materials

### Email to Scheduled Participants

Hello,

Thank you for agreeing to participate in the study "Jury Deliberation and the Not Criminally Responsible on Account of Mental Disorder Defence". We have scheduled you for [insert date and time].

You have been assigned as juror [juror number+jury letter]. Please bring this code with you to the study. You do not need to bring any other materials. Please aim to arrive 10 minutes early to sign in.

You will receive $40 in cash immediately upon completion of the study, in compensation for your time and for the cost of transportation (for parking or public transit, see below) to Carleton. We will go over the study procedure when you arrive at the lab, but we are happy to answer any questions you might have in advance. Below are directions to the Legal Decision-Making Lab.

Kind Regards,
Legal Decision-Making Lab

### Reminder email

Hello,

This is just a reminder that you are scheduled to participate in the study "Jury Deliberation and the Not Criminally Responsible on Account of Mental Disorder Defence" at 2111 Dunton Tower (21st floor), Carleton University, tomorrow, [insert date and time]. Please bring your assigned juror code with you to the lab.

If you have any questions, please don't hesitate to contact us. Thank you for your interest in our study.

Kind Regards,
Legal Decision-Making Lab

## Insufficient Number of Participants Script

Thank you very much for coming. Unfortunately, although we scheduled a number of other participants, too few people have attended today's session. In reality, a Canadian jury must have at least 10 members to continue, and so we cannot complete the experiment because we wouldn't be able to simulate a real deliberation closely enough. However, you will receive full compensation. I would be happy to answer any questions you might have and to tell you more about the purpose of our study. (See debrief script).

## Welcome and Instructions Script

Welcome and thank you for agreeing to participate in our study entitled "Jury Deliberation and the Not Criminally Responsible on Account of Mental Disorder Defence". Today, we are asking you to play the role of a juror. You will first read a short trial transcript (which will take about 15 minutes). We will then ask you to answer a brief questionnaire about what you have read. After that, we will ask you to begin group deliberation. Your task will be to come to a unanimous verdict decision. If after 45 minutes you cannot reach a verdict, then the jury will be declared "hung", a term that describes a jury that cannot come to a unanimous decision. After deliberation is complete, we will have a 5-minute debriefing session, during which I will provide you with additional information and answer any questions you might have. Please note that the correspondence associating your juror code with your email address will be deleted from our email account, and that information will be separated from our database as well.

There is a possibility you may feel uncomfortable with the nature of the trial transcript, which involves a second-degree murder charge and testimony about mental illness.

It is important for you to understand that this study will be audio and video recorded. We will also observe the study through this mirror. Your participation in this study is completely voluntary, and you have the right to refrain from answering any questions that you so choose. Your participation signifies that you consent to this, and that you understand that the video will be kept strictly confidential. You have the right to discontinue participation in this study at any point and you will still receive full compensation. If you no longer wish to participate you can simply exit the room and knock on the door of the room on the immediate right, where we will assist you promptly.

Once deliberations begin, although you may discontinue your involvement, we will retain the video footage for later analysis. These data will only be reported in aggregate form. Before the study begins, we will ask you to complete an informed consent form.

You will notice that you have 'name' cards with your juror code on them. These are for the purpose of communicating with each other during the deliberation and to protect your identity given that you will be video and audio recorded. We will now ask you to read this informed consent form that will further explain the purpose and procedures of this study. Once you have completed this task, please flip the paper face down.

## Informed Consent: Phase 2

The purpose of an informed consent is to make sure that you understand the purpose of the study and your involvement as a participant. The informed consent must include enough information regarding the study for you to be able to make a well-informed decision regarding whether or not you would like to partake in this study.

**Title**: Jury Deliberation and the Not Criminally Responsible on Account of Mental Disorder Defence

**Research Personnel:** This study is being conducted by Dr. Evelyn Maeder from the Institute of Criminology and Criminal Justice (evelyn.maeder@carleton.ca Tel. 613-520-2600 ext. 2421) at Carleton University in Ottawa, Ontario, Canada, and PhD student Susan Yamamoto (susan.yamamoto@carleton.ca) from the Department of Psychology.

**Concerns:** Should you have any ethical concerns regarding this study, please contact Dr. Shelley Brown (Chair, Carleton University Research Ethics Board – B, Shelley_brown@carleton.ca, 613-520-2600 ext. 1505). For other concerns, please contact the Research Compliance Office at ethics@carleton.ca.

**Purpose:** This is a study to evaluate opinions about the Not Criminally Responsible on Account of Mental Disorder Defence and about punishment in general.

**Task Requirements:**

You will first read a short trial transcript (which will take about 15 minutes). We will then ask you to answer a brief questionnaire about what you have read. After that, we will ask you to begin group deliberation. Your task will be to come to a unanimous verdict decision. If after 45 minutes you cannot reach a verdict, then the jury will be declared "hung", a term that describes a jury that cannot come to a unanimous decision. After deliberation is complete, we will have a 5-minute debriefing session, during which I will provide you with additional information and answer any questions you might have. Compensation for participating in both phases of the study is $40 in cash, which will be provided in person immediately following participation in Phase 2 at Carleton University. Full participation in Phase 2 of the study will take up to approximately 90 minutes.

**Potential Risk/Discomfort**: There is a possibility you may feel uncomfortable with the nature of the trial transcript, which involves a second-degree murder charge and testimony about mental illness.

**Right to Withdraw:** In addition, at any time, you may discontinue your involvement in this study without penalty. If you no longer wish to participate you can simply exit the room and knock on the door of the room on the immediate right, where we will assist you promptly. However, the video and audio data will be retained for later analysis, and so you cannot withdraw your data after you have already participated.

**Anonymity/Confidentiality:** The data collected are strictly confidential. Your name is not associated with the responses you provide. However, it is important to understand that you will be video/audio recorded, and these data will be retained indefinitely. In view of this we cannot absolutely guarantee the anonymity of your data. However, audio/video recorded deliberation data will not be stored online in order to reduce risk of data breach. Contact information will not be associated with these data in any way. The data will be kept in our lab facility on one password protected computer. Only coders/trusted colleagues will have access to these data. With your consent to participate in this study you acknowledge this. These data will only be used for research at Carleton University.

This study has been cleared by the Carleton University Carleton University Research Ethics Board – B (#16-023).

☐ *I have read the above form and understand the conditions of my participation. My participation in this study is voluntary, and I understand that if at any time I wish to leave the experiment, I may do so without having to give an explanation and with no penalty whatsoever. Furthermore, I am also aware that the data gathered in this study are confidential with respect to my personal identity.* ***By checking this box, I'm indicating that <u>I agree to participate in this study.</u>***

***OR***

☐ *I have read the above form and understand the conditions of my participation. My participation in this study is voluntary, and I understand that if at any time I wish to leave the experiment, I may do so without having to give an explanation and with no penalty whatsoever. Furthermore, I am also aware that the data gathered in this study are confidential with respect to my personal identity.* ***By checking this box, I'm indicating that <u>I disagree to participate in this study.</u>***

Print Name: _____

Signature: _____

I understand that I will be video and audio recorded, and that once group deliberation begins I will not be able to withdraw these data. I understand that I may discontinue participation at any time and still receive full compensation, and that these data will be kept strictly confidential.

**Please circle:**

I consent to be audio/video recorded
I do not consent to be audio/video recorded

Name:_____ Date:_____
Signature: _____

## B.1 Judge's Instructions

(Immediately prior to reading the case)

We will now ask you to read the trial transcript and answer the brief questionnaire. Once you have completed this task, please close your folder.

Gary Miller is charged with second degree murder. The charge reads: You must find Gary Miller not guilty of second degree murder unless the Crown has proven beyond a reasonable doubt that Miller is the person who committed the offence on the date and in the place described in the indictment. Specifically, the Crown must prove each of the following essential elements beyond a reasonable doubt:

1. That Gary Miller committed an unlawful act;
2. That Gary Miller's unlawful act caused Dennis Hughes's death; and
3. That Gary Miller had the intent required for murder.

If you are satisfied beyond a reasonable doubt of all these essential elements, you must find Gary Miller guilty of second degree murder.

The defendant, Gary Miller claims to have been mentally ill at the time that the crime charged in the indictment was allegedly committed. Since the law does not hold a person criminally accountable for his or her conduct while mentally ill (not able to appreciate), not criminally responsible on account of mental disorder can be used as a defence against the crime charged. The accountability of defendant Miller at the time of the alleged offence is, therefore, a question which you must decide.

To be found not criminally responsible, the defence must prove by clear and convincing evidence that:
1. Gary Miller had a severe mental disorder or defect at the time that the acts constituting the crime were committed, and
2. As a result of this severe mental disorder or defect, Gary Miller was not able to understand what he was doing or to understand that what he was doing was wrong.

The defendant must prove that he was mentally ill at the time of the offence by clear and convincing evidence; that is, the defendant must show that it is highly probable that he was mentally ill at that time.
In making your decision, you may consider evidence of the defendant's mental condition before or after the crime charged and you may consider not only the statements and opinions of any experts who have testified, but also all of the other evidence received in the case.
If you find that the defendant committed the acts described in the essential elements of Second Degree Murder, but that the defendant was not criminally responsible at the time that the acts were committed, you must find the defendant Not Criminally Responsible on account of Mental Disorder. Applying all of the other instructions given to you, you may find defendant Miller guilty if you decide Miller does not meet the requirements of Not

Criminally Responsible on account of Mental Disorder. Even though the defendant has raised the issue of mental illness, the Crown still has the burden of proving all the essential elements of the offence beyond a reasonable doubt.

(Immediately prior to deliberation)
I will now instruct you on the issue of mental disorder.

[1] You will consider this issue only if you are satisfied beyond a reasonable doubt that Gary Miller stabbed Denis Hughes, causing his death, and that he had the intent required for murder. If you are not satisfied beyond a reasonable doubt that Gary Miller committed the act, then he is entitled to an acquittal and there is no need to consider mental disorder.

[2]A person is not criminally responsible if he or she suffered from a mental disorder at the time of the act and, as a result, was not capable of appreciating the nature and quality of his act or of knowing the act was wrong. The issue has been raised whether Gary Miller is exempt from criminal responsibility on this basis.

[3] Every person is presumed not to suffer from a mental disorder. Exemption from criminal responsibility on this basis must be proved. I will now instruct you on a special rule on the burden of proof in relation to mental disorder.

[5] Gary Miller has raised the issue of mental disorder. You will recall that I earlier instructed you that the burden of proof beyond a reasonable doubt was on the Crown and that Gary Miller was not required to prove anything. The issue of mental disorder is an exception to this rule. Gary Miller must prove that it is more likely than not that he is exempt from criminal responsibility due to mental disorder at the time the offence was committed. This is a lower standard than proof beyond a reasonable doubt.

[6] To decide whether Gary Miller is exempt from criminal responsibility by reason of mental disorder, ask yourselves the following questions:

1. Is it more likely than not that Gary Miller was suffering from a mental disorder at the time of the act?

2. Is it more likely than not that Gary Miller's mental disorder made him incapable at the time either of (a) appreciating the nature and quality of the act or (b) knowing the act was wrong?

I will now review each of these questions with you.

[7] First - Is it more likely than not that Gary Miller was suffering from a mental disorder at the time he committed the act?

A mental disorder is a disease of the mind. A disease of the mind is any illness, disorder, or abnormal condition that impairs the human mind and its functioning. A disease of the

mind does not include a self-induced state caused by alcohol or drugs, or transitory mental states, such as hysteria or concussion.

This is the legal definition of disease of the mind and it is the definition you must apply, not any other definition that may have been used by counsel or the experts. I tell you as a matter of law that paranoid schizophrenia is a disease of the mind. It is for you to decide whether it is more likely than not that Gary Miller was suffering from paranoid schizophrenia at the time of the act.

If your answer to this question is yes, go on to the next question. If your answer is no, then Gary Miller is not exempt from criminal responsibility due to mental disorder.

Second - Is it more likely than not that Gary Miller's mental disorder made him incapable at the time either of (a) appreciating the nature and quality of the act or (b) knowing the act was wrong?

1. A person does not appreciate the nature and quality of an act if he does not know what he is doing, or does not foresee and understand the consequences of his act. The consequences refer only to the physical consequences, not the legal consequences.

2. Next, ask yourselves whether the mental disorder deprived Gary Miller of the capacity to decide rationally whether the act was wrong and, therefore, to make a rational choice about whether to do it. "Wrong" means morally wrong, judged by the everyday standard of the ordinary person. It does not mean legally wrong, and it does not mean wrong according to Gary Miller's own personal moral beliefs.

To reach the special verdict of not criminally responsible, it is not necessary that you find that Gary Miller was incapable by reason of mental disorder of both appreciating the nature and quality of his act and knowing that the act was wrong. As long as each of you finds that his mental disorder made him incapable of either one or the other, you do not have to agree on which one.

Ask yourselves whether it is more likely than not that Gary Miller's mental disorder at the time made him incapable either of (a) appreciating the nature and quality of the act or (b) incapable of knowing the act was wrong.

If the answer to this question is yes, then you must find Gary Miller not criminally responsible by reason of mental disorder.

[9] If you do not find that Gary Miller was suffering from a mental disorder, or if he was, it was not one that prevented him from appreciating the nature and quality of his act or from knowing it was wrong, you must still consider the evidence relating to mental disorder along with all the other evidence when you determine whether the Crown has proved the intent required for second degree murder beyond a reasonable doubt.

*Your verdict must be based solely on your consideration of all the evidence, the*

*submissions of counsel, and the law as I have explained it to you, and must not be influenced by a consideration of the consequences of a special verdict of not criminally responsible. However, you should know that our law provides that persons who are found not criminally responsible by reason of mental disorder will not be set free if they pose a significant threat to public safety.*

I will ask you to choose one juror to act as your foreperson. The foreperson will chair your discussions, and announce your verdict at the end of the case. Get to know each other a little before you choose your foreperson.

A verdict, whether of guilty, not guilty, or Not Criminally Responsible on Account of Mental Disorder, is the unanimous decision of the jury. To return a verdict on a count requires that all of you agree on your verdict. While your verdict on any count must be unanimous, your route to the verdict need not be.

You should make every reasonable effort, however, to reach a verdict. Consult with one another. Express your own views. Listen to the views of others. Discuss your differences with an open mind. Try your best to decide this case.

I am now handing you a verdict sheet. If you reach a unanimous verdict your foreperson should record it on your verdict sheet and notify me. We will come back to receive it.

If you cannot reach a unanimous verdict you should notify me in writing.

It is your duty to consult with one another and to try to reach a just verdict according to the law. Your foreperson will preside and assist you in the orderly discussion of the issues. You should each have the opportunity to express your own points of view without being unnecessarily repetitive. When you are discussing the issues, you should listen attentively to what your fellow jurors have to say. Approach your duties in a rational way and put your own points of view forward in a calm and reasonable manner. Avoid taking firm positions too early in your deliberations. Consider the views of your fellow jurors with an open mind before reaching your own decision.

We will not have a written transcript of the evidence available for you to review when you discuss your decision in this case. I think you will find that your collective memory of the evidence is good.

## B.2 Trial Transcript

*R v. Miller*

Alleged Crime: Second-degree murder
Victim: Dennis Hughes
Defendant: Gary Miller
D.O.B.: March 6th, 1989
Arrested: December 10th, 2014

## Crown Opening Statement

On the night of December 10th, 2014, Gary Miller stabbed and killed Denis Hughes. The facts meet the elements of second-degree murder; this is not in dispute by either side. The defendant further, by his own admission, knew that he was committing murder, and that murder is illegal. Thus, his action and admission satisfy all of the elements necessary to convict him of second-degree murder.

The reason we are here today is to determine whether the defendant is going to be held responsible for his actions. In pleading *Not Criminally Responsible on account of Mental Disorder*, the Defence actually assumes the burden of proof. It is true that the Crown always assumes the burden of proof in establishing guilt versus innocence, because the defendant is assumed to be innocent until proven guilty. However, the law has established that the defendant is assumed to be sane unless the Defence can prove otherwise. In other words, the burden for proving criminal responsibility rests with the Defence, not with the Crown. In this case, the Defence would have you believe that at the time of the offence, the defendant, Miller, was not capable of appreciating the quality of his act, of understanding that the act was morally wrong. But while you're listening to this testimony, I urge you to remember that the simplest explanation tends to be the right one. After you have seen the evidence you will understand that Mr. Miller is simply a violent person who snapped on his roommate during an argument that got so heated a neighbor overheard and called the police.

Although you will hear psychiatric testimony that, according to the words of the defendant, he was insane at the time of the offence, the Crown will show that actions speak louder than words in this case. You will hear testimony that the defendant stole money from the victim, and was preparing to skip town to avoid apprehension by the police. These are the actions of a person who knows he did something wrong but doesn't want to answer for it. Members of the jury, while Mr. Miller may have done an act that you and I believe only a sick individual could do, he was legally sane when he did so. Don't accept the Defence's farfetched fiction, the only evidence about which comes from a violent offender's claims, but instead hold him responsible. We ask you to return the only verdict appropriate in this case, Guilty.

**Defence Opening Statement**

Members of the jury, Gary Miller suffers from a very severe mental illness known as paranoid schizophrenia. The composed, seemingly rational person you see before you is a product of anti-psychotic medication. However, you have all heard the idiom, 'never judge a book by its cover.' To judge the inner workings of Mr. Miller's mind and mental illness based on his external appearance while he is on anti-psychotic medication is a grave mistake. You will hear medical testimony showing that the defendant has been positively diagnosed with paranoid schizophrenia.

Yes, it is true that Mr. Miller understood the legality of his act, but his actions were, in his mind, truly justified. The rationality of his belief, given his diagnosis of paranoid schizophrenia, is not the legal question to be decided here today. If a person kills another, and knows it is illegal but under an insane delusion that the salvation of the human race depends on it, then his action might be 'legally wrong', but it is not 'wrong' if we mean 'morally wrong'. The law is clear about this, members of the jury. Even if he knew it was against the law to kill another person, a mental disease caused him to think it was the only option. If you, the jury, feel that Mr. Miller, in his own mind, believed his actions were not morally wrong, then you must find that the defendant is Not Criminally Responsible on Account of Mental Disorder.

To return a verdict of Not Criminally Responsible you do not need to understand what he believed; surely, no sane person would believe that his loved ones have been replaced by alien imposters. What you do need to understand is *why* he believed what he did. The answer to that point is a severe mental illness or disease known as paranoid schizophrenia. It was only because of this mental illness that he stabbed Mr. Hughes. Keeping these facts in mind, the Defence calls on you to return a verdict of Not Criminally Responsible on account of Mental Disorder.

**Crown Witness, Officer Mark Hanes**

**Crown:** Can you please state your name and occupation?
**Hanes:** My name is Mark Hanes, and I am the police officer who was the first to arrive on scene on the day in question. I was also the officer that later arrested Gary Miller.

**Crown:** Can you please describe the events that took place on the evening of December 10th?
**Hanes:** At approximately 8:40 pm I responded to a 911 call from a neighbour about a disturbance at the apartment of Denis Hughes and Gary Miller. When I arrived, the door was open, and the victim, Mr. Hughes, was lying on the floor in the kitchen. I could see that the Mr. Hughes had lost a lot of blood, and I immediately called for medical assistance. Mr. Hughes was pronounced dead shortly after arriving at the hospital, and the cause of death was noted as multiple stab wounds to the neck and chest.

**Crown:** What happened next?
**Hanes:** We interviewed the neighbour who had called 911. We learned from this person that Mr. Hughes had a roommate, Gary Miller.

**Crown:** Was the defendant, Gary Miller, there at that time?
**Hanes:** No he was not.

**Crown:** When and where did you find Mr. Miller?
**Hanes:** At approximately 9:15 we found Mr. Miller at the home of his mother, Mrs. Miller.

**Crown:** Can you describe what happened next?
**Hanes:** We arrived at the home of Mrs. Miller, and identified ourselves as police officers. Mrs. Miller indicated that Mr. Miller was in his childhood bedroom. When we went to his room, it was evident the defendant was quickly attempting to pack some belongings. We let Mr. Miller know that we needed to ask him some questions pertaining to Denis Hughes.

**Crown:** Did the defendant comply with your instructions?
**Hanes:** No he did not. Mr. Miller attempted to flee through the bedroom window, which was on the ground level of the house. At that time we apprehended Mr. Miller and took him in for questioning.

**Crown:** Did you find anything of note at Mrs. Miller's home?
**Hanes:** Yes. We found a butcher's knife, which we later identified as the murder weapon. Mr. Miller had cleaned the knife in the bathroom sink. We also recovered a wallet, which contained $200 cash and several cards; it belonged to the victim, Mr. Hughes.

**Crown:** So, to summarize, Mr. Miller was attempting to pack his belongings, and to

avoid capture?
**Hanes:** Yes.

**Defence Cross-examination:**

**Defence:** What was Mr. Miller's demeanor at the time you arrived at his house?
**Hanes:** He seemed frantic, and unsettled by sudden police presence.

**Defence:** Did he say anything?
**Hanes:** He shouted something to the effect of: "Get away from me, don't let them take me".

**Defence Witness, Dr. Devin Cassady**

**Defence:** Can you please state your name and occupation for the court?
**Cassady:** I'm Dr. Devin Cassady. I'm a psychiatrist working at the Forensic Psychiatric Institute.

**Defence:** What are your credentials?
**Cassady:** I am a medical doctor, and a Fellow of the Royal College of Physicians. I've been practicing psychiatry for over 20 years now.

**Defence:** Have you spoken extensively with the defendant, Gary Miller?
**Cassady:** Yes. I conducted a full psychiatric assessment of Mr. Miller.

**Defence:** What did you learn from this assessment?
**Cassady:** Based on a psychiatric and medical history, a standardized questionnaire, and my own more detailed interview, it is my professional opinion that Mr. Miller meets the diagnostic criteria for schizophrenia, paranoid type.

**Defence:** Can you describe for the courts what exactly 'schizophrenia' is?
**Cassady:** Schizophrenia is a severe brain disorder in which people interpret reality abnormally. Schizophrenia may result in some combination of hallucinations, delusions, and extremely disordered thinking and behaviour. Its origins are not yet fully understood by scientists, but its potential debilitating effects are well documented.

**Defence:** Could you please tell the jury some details about how someone is diagnosed with schizophrenia, and what that means?
**Cassady:** Well, first you must rule out other mental health disorders and determine that the symptoms aren't due to substance abuse, medication, or a medical condition. In addition, a person must have at least two of a specific set of symptoms outlined in the Diagnostic and Statistical Manual of Mental Disorder (also called the 'DSM'), and those symptoms would be present for most of the time during a one-month period, with some level of disturbance being present over six months. We look for things like delusions, hallucinations, disorganized speech (indicating disorganized thinking), and extremely disorganized behaviour.

**Defence:** What did you learn, during your assessment, about Mr. Miller's behaviour on the day in question?
**Cassady:** Mr. Miller suffers from what is called "Capgras Delusion", which can occur in patients with paranoid schizophrenia. The key feature of this delusion is that the patient believes that his loved ones have been replaced by identical looking imposters. Mr. Miller indicated to me that he believed that an alien imposter, who had transplanted a chip into his brain, had replaced his roommate. This chip, he believed, was responsible for his hearing of Mr. Hughes's voice even when he was not present. He told me he suspected that aliens were conspiring to take over the planet, and that the Mr. Hughes imposter was attempting to extract information from his mind. He remarked to me that he began to suspect this was the case a couple of months prior, when he came home to find

that Mr. Hughes had moved the TV to a different spot in the room.

**Defence:** In your discussions with Mr. Miller about the night in question, what did he tell you?
**Cassady:** He recalled that he and Mr. Hughes were talking in the kitchen, and that he heard a knock on the door. He believed that Mr. Hughes intended to take him away to a secret facility that night, and that he had to kill him to get away.

**Defence:** In your opinion, is Mr. Miller trying to mislead you into believing he has schizophrenia?
**Cassady:** No I do not. Mr. Miller presented with classic symptoms of schizophrenia, and in particular, Capgras delusion.

**Crown Cross-Examination**

**Crown:** Are you an expert in deception, Dr. Cassady?
**Cassady:** No, I am not. But I have many years experience treating real illnesses, and the ability to detect malingering can be part of the job.

**Crown:** Is that because people sometimes lie, and try to trick their doctor into diagnosing them with an illness?
**Cassady:** It's a possibility, but to my knowledge it is not all that common. Capgras Delusion specifically is not necessarily well known to laypeople, so it wouldn't really be something one would fabricate easily.

**Crown:** But, wasn't your diagnosis just based on what the defendant told you, after some time had passed following the incident?
**Cassady:** I conducted a full psychiatric assessment. From this assessment, it is my opinion, based on many years of experience, that Mr. Miller's behaviour is consistent with a diagnosis of schizophrenia.

**Crown:** But in fact, you didn't even interview Mr. Miller until two full weeks after the crime had occurred, isn't that right?
**Cassady:** Yes that's correct.

**Crown:** Do you think that is enough time for someone to research the symptoms of schizophrenia, or research the insanity defence?
**Cassady:** I don't really know – I couldn't speak to the defendant's activities during that time.

**Crown:** Yet you can be confident about his mental state at the time of the crime? That he has this bizarre, specific delusion based only on a description of his activities?
**Cassady:** I conducted a full psychiatric assessment. That means that I had to take into account a lot of factors, not just Mr. Miller's word. I looked for things like certain speech patterns, emotional expression, thinking and perception spanning the months leading up to the incident and at the time of assessment. We're not just looking for what patients say,

but how they behave – it's not as simple as just making up stories.

**Crown:** So, you're saying that the victim believed aliens replaced his roommate and he had to escape quickly, but he still felt he had time to take the victim's wallet? Was that part of this 'delusion' as well?
**Cassady:** The point is that Mr. Miller's behaviours were erratic, frantic, because he was under the influence of paranoid delusions. For example, he indicated that if the police captured him, the aliens could get to him easily. Although the behaviour is irrational to a person who is well, it is reasonable to suspect that taking the wallet somehow played into those delusions.

**Crown:** Speaking of psychiatric history, to your knowledge, has the defendant ever been hospitalized for paranoid delusions before the incident?
**Cassady:** No he has not.
**Crown:** Thank you Dr. Cassady, that's all I have for you today.

**Defence Re-direct**

**Defence:** Dr. Cassady is it surprising to you that Mr. Miller would not have spent time in a psychiatric facility?
**Cassady:** Not necessarily. Among men, onset of schizophrenia typically occurs during early to mid 20's. Even if Mr. Miller began experiencing disturbances before the incident, he would not likely have understood the need to seek treatment. Having moved out of his family home, his family would not per se be able to intervene either.

**Crown Closing Statement**

Mr. Hughes woke up on December 10th, excited to finish his last exam before the holidays and soon go home to friends and family. He was a good student, and had many exciting plans in store, but instead, his life was cut short. I would like to remind you, ladies and gentlemen, that Mr. Miller does not deny intentionally ending Mr. Hughes's life. So, you don't need to take my word for it, instead you can take this information directly from the defendant. The disturbing truth is that Mr. Miller is a dangerous, cold-blooded killer. Frustrated with his roommate, he snapped and violently silenced Mr. Hughes. He knew that it was illegal, and he knew he would get in trouble. We can clearly see this because he took some quick cash from the victim, fled the scene, and even cleaned the murder weapon. Once he got caught red handed he had to come up with a good story. Members of the jury, do not fall for his fanciful story. What is likely: that the defendant suddenly experienced paranoid delusions even though we have no evidence of this? The defence is so insistent that these outlandish beliefs excuse Mr. Miller from criminal responsibility, and yet you didn't hear from a single witness who could attest to any strange behviour in the months leading up to the crime. Or, is it more plausible that he is just a violent person who lost his cool when he argued with his roommate one too many times? Not one piece of evidence, besides the defendant's own account, that he had a mental disease at the time of the event. What we do have is overwhelming evidence of second-degree murder: the body of Mr. Hughes, a murder weapon, the defendant's belongings in Mr. Miller's room, and even a direct admission. While the psychiatrist you heard from might not be able to tell when someone is faking an illness, I have every confidence that you can, members of the jury, and that you will return the correct verdict in this case: Guilty.

**Defence Closing Statement**

This is a very tragic case, ladies and gentlemen; there is no doubt that. Mr. Miller also deeply feels the loss of his best friend and roommate, Mr. Hughes. The real culprit here is mental illness. We have shown you beyond a shadow of a doubt that Mr. Miller could not appreciate nature and quality of his act. You heard testimony from a very experienced doctor describing an undisputed diagnosis of paranoid schizophrenia. The Crown would have you believe that because Mr. Miller's account is so bizarre, it can only be a piece of fiction. But after hearing Dr. Cassady's testimony, you can understand that the reason these beliefs sound so far-fetched to you or me is because they came from a mind that is unwell. Don't fall into the trap of attempting to understand Mr. Miller's delusion. The Crown is trying to distract you from the real legal issue at hand: namely, whether Mr. Miller believed these things because of a mental illness. Trust in an expert's full assessment that was based on a lot more information than you've heard in this case. It was based on years of training, experience, and study of mental health. Let us not punish Mr. Miller for being the unlucky recipient of a mental disease that consumed his life and left him in fear for it. What Mr. Miller really needs to receive is psychiatric care from trained medical professionals who understand how the brain works. I trust you, members of the jury, to follow the law in this case. The law tells us that if a person could not appreciate the nature and quality of his act, then he will not be held criminally responsible, plain and simple. This doesn't mean he can just walk out of here, it just means that we can't blame him for something that was so beyond his control. This isn't a case of evil, but rather it is a case of illness. While Mr. Miller did not have the choice to act rationally at the time of the crime, you have a choice here and now; the rational one is to find the defendant Not Criminally Responsible on Account of Mental Disorder.

1. How do you find the defendant?

- Guilty
- Not Criminally Responsible on Account of Mental Disorder

2. How confident do you feel in your verdict?

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Not at all confident | | | | | | | | | | Very confident |

**Deliberation Instructions Script**

We will now ask you to begin your deliberation. Again, if you no longer wish to participate at any point, or if you require assistance, you can simply exit the room and knock on the door to the immediate right. We kindly ask that you refrain from use of cell phones throughout the deliberation except in case of an emergency. I will come back in 45 minutes.

(After 45 minutes)

Were you able to reach a unanimous verdict?
a. (If yes): How do you find the defendant?

b. (If no): In the event that a jury is unable to reach a unanimous verdict, the court usually declares a mistrial.

(Begin debriefing script)

**Debriefing Session Script**

Thank you all for your participation. This debriefing session will clarify the purpose of our study and why we are interested in this issue. The purpose of this study was to understand how jurors process and discuss information in a Not Criminally Responsible on Account of Mental Disorder trial.

The trial transcript that you read was not an actual case, but instead was fabricated for the purpose of this study and modeled after several different cases. We were not able to inform you that this was a fictional case at the outset, because participants are sometimes more unbiased in their judgment if they do not believe that they are reading about a real victim and defendant, whereas we want to simulate as best as possible true jury decision-making.

What were we trying to learn in this study? Previous research has shown that jurors have somewhat negative views of the insanity defence, which might mean that they are hesitant to apply it, even in appropriate cases. However, because this trial was fabricated, there is no right or wrong verdict decision. Rather, we are interested in what types of information people think is relevant to discuss during deliberation. To our knowledge, this is the first NCRMD deliberation study in Canada, and such a study can give us insight into how Canadians feel about the insanity defence and when they think it can be used.

We do not have specific predictions about your discussion or verdict decision, but we will look at whether there are any themes that emerge across 15 different mock juries. If you would like to receive an email summarizing these results, please write your email address on this sheet.

If you have any questions, I can answer them now, or you can email me questions any time. You may take this contact information home with you in case you have any questions about this study, about ethical concerns, or any other concerns.

The study is now complete. Thank you very much for participating; we value your time and opinions.

(Provide compensation and contact information sheet).

**Appendix C Codebooks**

| Category | Subcategory | Additional Categories | Operational Definition |
|---|---|---|---|
| **Injustice/Danger** | Expresses insanity myth | Returned to street | Believes NCR defendants are normally released prematurely |
| | | Insanity easily faked | Believes mental illness is commonly faked or easily faked (e.g., mental illness assessments are not multifaceted) |
| | | NCR frequently used | Believes insanity pleas are a common 'fallback' |
| | Corrects insanity myth | Returned to street | |
| | | Insanity easily faked | |
| | | NCR frequently used | |
| **Strict Liability** | Strict Liability | | Mental illness is irrelevant to criminal responsibility ("do the crime, do the time") |
| **Utilitarianism** | Deterrence | | The notion that we can stop future crimes (for other or for the same individual) via punishment |
| | Incapacitation | | The need to restrict an offender's ability to perform an act; to simply stop the behaviour (e.g., imprisonment, exile) |
| | Rehabilitation | | The need to morally educate or treat an offender with the ultimate goal of wellness and/or re-entrance into society |
| | Greater good | | It is necessary to maximize happiness and minimize suffering for the greatest amount of people, even at the expense of one person's rights |
| | Benefit society | | Punishment should have a future benefit in order to be justified (i.e., we cannot punish for the sake of punishment alone) |
| **Retributivism** | Prefer guilty | | It is not permissible to punish people who are not |

| | | | |
|---|---|---|---|
| | unpunished | | responsible; prefer to protect individuals from excessive or unjust punishment |
| | Just deserts | | The notion that punishment is necessary to balance the scales of justice, or that justice needs to be done for the victim |
| | Criminal blame | | Criminals bring punishment on themselves (i.e., they have control over their circumstances) |
| **Case Narratives** | Believing | | Indicates that Crown's story more believable |
| | Doubting | | Indicates that Defence's story more believable, or Crown's story is not believable |
| | Mulling | | Indicates reasons why both stories are plausible |
| | Puzzling | | Unable to form a story; indicates that the case is confusing |
| **Scope** | Rich | | Indicates that there is enough evidence in the case |
| | Lacking | | Indicates that there is not enough evidence in the case |
| **Credibility** | Psychiatrist | Credible | |
| | | Not credible | |
| | Police officer | Credible | |
| | | Not credible | |
| | Defendant | Credible | |
| | | Not credible | |
| | Crown Counsel | Credible | |
| | | Not credible | |
| | Defence Counsel | Credible | |
| | | Not credible | |
| **Story Sampling** | Isolated statement | | Makes brief hypothetical about singular piece of information without putting it into full context. |
| | Complete story | | Gives full hypothetical about the event. |

| | | | |
|---|---|---|---|
| **Case Evidence** | Wallet | Defence | Cites testimony that the defendant was in possession of deceased's wallet in favour of NCR. |
| | | Prosecution | Cites testimony that the defendant was in possession of deceased's wallet in favour of guilt. |
| | "Don't let them take me" | Defence | Cites testimony that the defendant uttered this phrase in favour of NCR |
| | | Prosecution | Cites testimony that the defendant uttered this phrase in favour of guilt |
| | Psychiatric testimony | Skepticism | Doubts psychiatrist's testimony |
| | | Trust | Believes the psychiatrist is qualified, had sufficient information |
| | Fled the scene | Defence | Cites testimony that the defendant uttered this phrase in favour of NCR |
| | | Prosecution | Cites testimony that the defendant uttered this phrase in favour of guilt |
| | Mental illness history | Defence | Cites psychiatric testimony that defendant was average age of onset |
| | | Prosecution | Cites lack of mental illness history in favour of guilt. |
| | Medication | | Refers to testimony that defendant is receiving anti-psychotic medication |
| | Time delay of assessment | | Cites testimony that assessment delay allows for time to fabricate insanity story. |
| | Other | | In vivo (quote directly) |
| **Group-related** | Polling | Simultaneous | Jurors publically declare their vote at the same time |
| | | Sequential | Jurors publically declare their vote in turn |
| | | Interrupted | The poll is interrupted by discussion |

| | | | |
|---|---|---|---|
| | Yielding | To NCR | Juror changes position to NCR |
| | | To guilt | Juror changes position to guilty |
| | Championing | NCR Champion | Juror advocates for NCR |
| | | Guilty Champion | Juror advocates for guilty |
| | Holdout | | Juror maintains position even when all other parties are in favour of opposition |
| **Legal Instructions** | Reasonable doubt | Of guilt | References notion that there is reasonable doubt |
| | | Of mental illness | References notion that there is reasonable doubt |
| | Burden of Proof | Defence hasn't met | Asserts that defence hasn't met burden of proving that defendant has mental illness or that a mental illness precluded a guilty mind |
| | | Crown hasn't met | Asserts that crown hasn't met burden of proving the defendant's guilt (e.g., motive) |
| | Disposition | Can't consider punishment | References instruction not to think about the disposition when coming to a decision |
| | | Won't go free instruction | References instruction that defendant's premature release is not an issue |
| **Emotions** | Intuition | | Does not articulate rationale, or instead refers to "feeling" |
| | Anger at group | | Chastises or criticizes group. Raises voice aggressively |
| | Anger at defendant | | |
| | Sympathy for defendant | | |
| **Strength of Conviction** | Confident | | Expresses certainty in position (e.g., "this case is clear cut") |
| **Suspension of disbelief** | Case is fictional | | Fixates on fictional nature of trial; indicates that the story is too strange to be real |
| | Social experiment | | Discusses the fact that the deliberation is a social experiment; speculates about the research team's motives |

| Keyword /Content | Exemplar | Code | Category |
|---|---|---|---|
| **Evidence** | "Yeah, but who said that "Don't let them take me"? He did" | Question | **Established evidence** |
| | "Yeah, took a wallet with money in it." | Fact in evidence | |
| | "took money because you know he realizes you know he's he even if he uh you know he does, like he could be you know the aliens are still gonna come after him cuz he just killed his roommate, I think" | Defendant's thinking | **Simulation** |
| | "A sane person, if they killed someone, would get rid of the weapon and not take it to a place that [gestures to head] you would assume...," | Sane people | |
| | "I feel like, he didn't react like someone who was following his own delusions, he acted like someone who was like *Oh crap! I just killed someone, that's bad!*" | I feel like | |
| | "We don't know that, can just speculate on that." | Shouldn't speculate | |
| | "I think him cleaning the knife especially – if he was really concerned that an alien was trying to kill him or something…he's not going to take the knife or take the wallet of what he thought was this other person right..." | If/then | |
| | "From the case, I would like there to have been more outside opinion, I think the history of violence is something that plays into it." | Desired evidence | **Deficient** |
| | "We have to remember there wasn't a crown attorney appointed psychiatrist too." | Missing evidence | |
| | "There's so many things missing!" | Scope | |
| | "The lack of detail bothers me." | Lack of detail suspicious | |
| | "The police are sticklers for detail, if he was doing that he would have said so" | Lack of detail | |

| | | | |
|---|---|---|---|
| | "And we had no witness prior to the uh, to the court case of anybody saying he has abnormal behaviors" | No witnesses | |
| | It means that they have a weak case and they know it. | Weak case | |
| | "There are many criteria that have to be met, and many tests that are used, depending on where you are." | | **Sufficient** |
| | "But still, I mean I think the Crown brought up, valid objections, like the fact that there was a gap, umm, the fact, you know, there wasn't uh, you know, th-there was a brief mention of-of his demeanour when the cops arrived, but no other like, no nothing from even like, his mother, whose house he went to, like" | | |
| | "He's given us too much to be able to give him the benefit of the doubt" | | |
| **Disposition** | "Does he kill again? Because he… he's [scare quotes] schizophrenic?" | Kill again | **Future harm** |
| | "I, so then Garry Miller get's to go free?" | Goes free | |
| | "He is a danger to society." | Dangerous | |
| | "Either way, he should be like, confined, either just like, as a prisoner criminally, or as a patient." | Confined | |
| | "That's a little bit out of our jurisdiction, that would be more up to the judge based on whatever the judge decides based on that." | Not our jurisdiction | **Not our duty** |
| | "It's that or spend the rest of your life in prison." | Alternatives | |
| | "The thing is, you can wind up spending longer in psychiatric facility on the not guilty, the not criminally responsible than you would in a jail." | Release timing | **Release** |

| | | | |
|---|---|---|---|
| | "You send him to jail, he's gonna do his time, he's gonna get out, and he'll be back in fruitbat land." | Release timing | |
| | "The one thing that's guaranteed though if you're sent to the psychiatric facility, he will be treated. Okay?" | Guaranteed treatment | **Rehabilitation** |
| | "Well, I agree with you the prison system does nothing to rehabilitate criminals." | Prison not rehabilitative | |
| | "Not that he's not going to go to, y'know, psychiatric hospital or have punishment, he's gonna go to the doctors rather than you know, just a general prison." | Have punishment, psychiatric hospital | |
| | "It's not whether we feel for the, the victim, it's a legal tragedy if, if, if uh, a mentally person, troubled person gets uh, goes to jail." | Not about feelings/Legal tragedy | |
| | "So he's still not getting off scot-free, not criminally responsible (inaudible) so" | Not scot-free | |
| | "No, but yeah, we can't decide the sentence" | Can't decide sentence | |
| | "I do agree, like I do agree like it's not our decision to judge but it's still…" | Not our decision | |
| | "I know it's a hard call, but that's why I asked… are we allowed to do that?" | Are we allowed? | |
| | "Yea, but I feel like that would have less long-term consequences?" | Long-term consequences | |
| | "There's a whole, there's a whole thing of, I-I mean, I'm personally, strongly, of the opinion that the prison system needs reform, but like, I still, I mean…" | Prisons need reform | |
| | "I would just have to hope that, if someone were sent to prison, and they clearly have a mental disorder? [Shrugs shoulders]" | Type B error | |
| | "if he doesn't need that treatment, that's a lot of, of-of resources that someone else could really need, so it's just like…" | Wasted resources | |

| | | | |
|---|---|---|---|
| **Moral Foundations** | | | |
| **Authority** | | | |
| "Law" | "It's just, that is the law." | That's the law | **Deference to authority** |
| "Law" | "That's like what you want for him but that's not the law." | Not the law | |
| "Authority" | "I think you should take the authority of that, uh, that worker."<br><br>"The doctor's uh, opinion seemed fairly authoritative" | Defer to expert | |
| "Mother" | "the fact that we didn't hear from the mother might signal that she's trying to protect him" | Mothers protect | |
| "Mother" | For me, actually my personal experience with it was that I was raised by my mother who had paranoid schizophrenia | Anecdotes | **My credentials** |
| "Law" | "I worked in law firms for over six years" | My experience | |
| "Position" | "I mean, I'm not the doctor so I can't put myself in the position of what a schizophrenic would be thinking, feeling, going through all these wild thoughts." | Not a doctor | |
| **Fairness** | | | |
| "Reasonable" | "I don't believe the Defence has, proved to insanity to a reasonable degree." | | |
| | "I found explanations were reasonable." | | |
| "Bias" | "And even though he has twenty years experience he's still gonna, to have, um, bias, some personal bias." | Doctor biased | |
| | "I think also the doctor who like you said [points to J1] has no particular bias in this issue." | Doctor unbiased | |
| | "Yeah, I'm, um, again, I'm totally biased as well" | I'm biased | |
| "Prejudice" | "I mean, his mother would be prejudiced but, like other people." | | |
| "Justice" | "But that's kind of like a flaw of the justice system" | | |

| | |
|---|---|
| "Justified" | "…in law under no circumstances are you justified in using deadly force, even in your own self-defense." |
| | "What justifies killing an alien?" |
| "Fair" | "Which is fair." |
| | "…to be fair, what if he has researched it?" |
| **Harm** | |
| "Kill" | "He killed a man" |
| | "Like ninety percent of people who kill people have a history of violence." |
| | "People being rational, happy, healthy, well-developed people don't kill their roommates." |
| | "Yea I think, I think he knew that he killed him" |
| | "Yeah I don't think you would kill a guy for two hundred bucks." |
| "Fight" | "but you're fighting for your life –" |
| | "And there's no recollection of him fighting with the police to get away." |
| "Suffer" | "There's nowhere it even tells me that he's actively suffering, currently." |
| | "I think most of us agree, that he is suffering from what he says he's suffering." |
| "Wound" | "One thing that we haven't talked about is the wounds themselves." |
| "Abuse" | "I mean most of them either abuse animals before, or get in fights and assaults before the murder, they don't just –" |
| "Violent" | "He would have done something violent beforehand." |
| "War" | "It's like going to war. We all agree that war – that murder is wrong, except in war then it's okay to kill people right." |

| | |
|---|---|
| "Protect" | "he had to protect… society by killing this guy and he did what he did…" |
| | "I think that if he was schizophrenic the, uh, the overriding moral priority of protecting, the planet?" |
| "Benefit" | "so I'm just saying if it's to the point where we're really indecisive? Then we should go with the decision which, is the most beneficial, or protecting of, people, including, inmates, but, I, just [raises hands] like." |
| "Safe" | "The administrator can decide release when they decide he's safe…" |
| | "I guess people usually seek somewhere that they feel safe" |
| "Sympath" | "I think we were sympathetic to your side." |
| | "They're very lefty people, they raised me very much to the left and between them and the Johnny Cash and more sympathy for criminals than…" |
| | "I think you guys are so, unsympathetic, and you're so young too (inaudible)." |