Running Head: The Bastard Verdict.

Title: The Bastard verdict.

Lee J. Curley[1*], Rory MacLean[1], Jennifer Murray[2], Phyllis Laybourn[1], & David Brown[3]

1. School of Applied Sciences, Psychology Subject Group, Edinburgh Napier University, Edinburgh, Scotland, EH10 4BN.
2. School of Health and Social Care, Mental Health Theme.
3. Currently not affiliated with any Universities.

Word count (exc. tables/references): 5682.

*Requests for reprints should be addressed to Lee Curley, School of Applied Sciences, Psychology Subject Group, Edinburgh Napier University, Edinburgh, Scotland, EH10 4BN. (email: Lee.Curley@napier.ac.uk).

Abstract

The Scottish Legal system is a unique jurisdiction as jurors are able to give Not Proven verdicts in addition to the well-known Anglo-American verdicts (Guilty and Not Guilty). The Not Proven verdict has never been legally defined, meaning that currently legal practitioners can only estimate why a Not Proven verdict has been given. The main aim of this study was to investigate if jurors violate the regularity principle, which is commonly incorporated in many rational choice models, by testing if the introduction of the Not Proven verdict has an impact on the outcomes given by jurors. In addition, this study aims to test if the introduction of the Not Proven verdict has an impact upon how the Not Guilty verdict is perceived by jurors. In this study, 128 participants listened to two vignettes centred on homicide trials, jurors could give one of two verdicts in one of the vignettes and one of three verdicts in the other vignette. The vignettes were counterbalanced in regard to how many verdicts could be given at the end of them. It was found that jurors in a three-verdict system were less likely to give a Not Guilty verdict in comparison to jurors in a two-verdict system, showing that jurors violate the regularity principle and that the Not Proven verdict may change how the Not Guilty verdict is perceived. The current paper has implications in relation to juror communication, article six of the European convention of human rights and juror rationality.

Keywords: Law, Not Proven verdict, Human rights, Rational decision making, Psychology.

Acknowledgements: This research was carried out in partial fulfilment of the first author's PhD. The Mcglashan Charitable trust provided the lead researcher with funding to help pay tuition fees when this research was being conducted. The study was awarded ethical approval from Edinburgh Napier University's School of Applied Sciences' Research Integrity Committee in 2017. The authors have no competing interests to declare.

Guarantor: Lee Curley will act as the guarantor for this work, accepting responsibility for the conduct of the study, holds access to the data, and controlled the decision to publish, though

all study authors were involved in the preparation of the manuscript, interpretation of the data, had access to the data, and informed the study design.

Statement of Contributions: The conceptualisation for the study initially arose from Lee Curley, with the remaining authors assisting in shaping the specific aim and design of the research. Lee Curley collected and analysed the data. Lee Curley, Rory MacLean and Jennifer Murray interpreted the data. Lee Curley prepared the first and final drafts for publication; Jennifer Murray, Rory MacLean and Phyllis Laybourn assisted in preparing drafts of the manuscript. Mr David Brown helped design the researching materials, and was crucial in allowing the study to be conducted.

The Bastard verdict and its influence on jurors.

Introduction

Development of the Not Proven verdict.

The Scottish legal system is very different from other jurisdictions [1]. For instance, within the Anglo-American legal system only the verdicts of Guilty and Not Guilty are available [2]. Therefore, if the evidence does not provide proof beyond reasonable doubt that the defendant committed the accused illegal act, then a Not Guilty verdict is appropriate [2]. However, within the Scottish legal system, three-verdicts are available: Guilty, Not Guilty, and Not Proven. To understand how this system developed, we need to look back at the late 17th century and early 18th century [3]. In this time period, jurors where only given the opportunity to make a decision on whether the charges were not proven or proven based on the evidence, and the judge decided on the guilt of the suspect [3]. This system was introduced because from 1660 onwards juries declined to give Guilty verdicts on oppressive laws [3], thus the introduction of this system was to attenuate the power of the jury. In the early 18th century, the ability of the jury to decide on the guilt of the defendant (i.e., give a Guilty or Not Guilty verdict) was re-established [3]. However, the Not Proven verdict remained, and consequently the Scottish three-verdict system was born.

Scottish law does not define what a Not Proven verdict is, and judges have been deterred from helping jurors to interpret said verdict [4]. It is believed, however, that the Not Proven verdict symbolises that proof of guilt has not been proven, whereas a Not Guilty verdict in the Scottish courtroom shows that jurors believe the suspect to be truly innocent [4]; the current study aims to test these assumptions. Further, from a legal perspective, the Not Proven verdict has the same outcomes as the Not Guilty verdict (i.e., no custodial sentencing, the defendant is discharged and not subject to any sentence).

Review of Empirical studies surrounding the Not Proven verdict.

Hope, Greene, Memon, Gavisk, and Houston [5] found that jurors in a three-verdict system (i.e., Not Proven, Guilty, and Not Guilty) were less likely to give a Not Guilty verdict in comparison to jurors in a two-verdict system (i.e., Guilty or Not Guilty). The estimate of guilt given for Not Proven verdicts, however, was not significantly higher (Not Proven M = 52.4%) than the estimates given for Not Guilty verdicts in either of the verdict systems (Not Guilty, two-verdict M = 42.4%; and, three-verdict M = 42.5%). This highlights that the Not Proven estimate of guilt given in Hope et al.'s [5] research was large enough to justify giving a Not Proven verdict over a Not Guilty verdict, but was not large enough to allow a Guilty verdict to be given.

In addition, Hope et al. [5] demonstrated that irrespective of verdict or evidence weight, verdicts made in the three-verdict system allowed the mock jurors to feel more confident that they had reached the correct verdict than their binary verdict counterparts. Confidence does not always equal accuracy, however (Hall et al., 2007). Hope et al. [5] also found that the different verdict systems did not have a significant impact on the juror's perception of guilt surrounding the defendant.

Bhatia [6] found that if a third response is available which supports one of the original responses but not the other, then the option supported by the additional third response will be the response most likely to be chosen. For example, if one group of participants are asked which packet of crisps contained the highest fat content between a 150g bag of Walkers, a 150g bag of Golden Wonder, or a 100g bag of Walkers, and a second group is asked which packet of crisps contained the highest fat content between a 150g bag of Walkers and a 150g bag of Golden Wonder, the first group would be more likely to choose the 150g bag of Walkers in comparison to the second group. Bhatia [6] has found this result consistently using his own tests and examples.

Bhatia [6] suggested that asymmetric dominance could explain these results. Asymmetric dominance essentially means that if an additional third option is added, which is recessive in one factor to one of the original options but similar to that same option in regard to another factor, and different to an alternative option, it will allow the similar yet dominant option to be considered the leading choice. In the crisp example above, the 150g bag of Walkers has a higher chance of being picked in the three-option condition in comparison to the two-option condition, as the additional third choice is made by the same company but is recessive in terms of weight. This research showed that the introduction of an extra option could cause differences in relation to which option is picked most frequently.

If Bhatia's [6] and Hope et al.'s [5] findings are taken together, an interesting conclusion can be drawn: that the introduction of a third option influences which option, out of an original binary set of options, is mostly likely to be chosen. Asymmetrical dominance, however, cannot explain the results found in Hope et al.'s [5] study, as the Not Proven verdict is similar to the Not Guilty verdict as both are acquittal options, but is also recessive to the Not Guilty verdict because it is not an outright acquittal. Nevertheless, the frequency of Not Guilty verdicts did not increase in the three-verdict system in comparison to the two-verdict system, thus showing that asymmetrical dominance may not fit in well within a juror setting.

Smithson, Deady, Gracik [7] conducted a similar experiment to Hope et al. [5]. Their first study investigated how the introduction of a third verdict may affect the verdict chosen, the belief of guilt and the difficulty of the decision in both a civil case surrounding negligence and a criminal trial focussed on murder. A significant association between the verdict that was given and the number of verdicts that were available was found, fewer Not Guilty and Guilty verdicts (to a lesser extent) were given in the three-verdict condition in comparison to the two-verdict condition; and, when the Not Proven verdict was available, significantly more guilty verdicts were given in comparison to not guilty verdicts [7]. Once again, this suggests

that asymmetric dominance may not apply to the courtroom. Middling beliefs of guilt led to Not Proven verdicts, causing the beliefs of guilt associated with Not Guilty verdicts to be reduced in both trial types [7].

Smithson et al.'s [7] second study replicated the study outlined above, but added guilty for Manslaughter as an additional potential verdict in one of the conditions. Within the study, there were three conditions: 1) Anglo-American two-verdict system (only guilty for murder verdict and not guilty verdict available); 2) the Scottish three-verdict system; and 3) a three-verdict system where the third verdict was the guilty for Manslaughter verdict [7]. It must be mentioned here that the guilty for Manslaughter verdict is not directly comparable with the Not Proven verdict, as the guilty for Manslaughter verdict may be used when the jury believes that the defendant caused the death of the victim but indent was not there; it is a conviction but for a lesser offence. It was found that Not Proven and guilty for Manslaughter verdicts seemed to push participants away from giving Not Guilty and guilty for murder verdicts, and both Not Proven and guilty for Manslaughter verdicts did not differ significantly in terms of similarity ratings when contrasted with guilty for murder verdicts. Their third study focussed on the similarities between each of the verdicts (guilty for murder, Not Guilty, guilty for Manslaughter and Not Proven), and found that guilty for Manslaughter verdicts were more similar to guilty for murder verdicts and that Not Proven verdicts were more similar to Not Guilty verdicts [7]. In addition, Smithson et al.[5] found that guilty for Manslaughter and guilty for murder verdicts did not differ significantly in relation to belief of guilt ratings. Not Proven verdicts had a significantly lower belief of guilt than Guilty verdicts, and Not Guilty verdicts gave a significantly lower belief of guilt than Not Proven verdicts. Decisions that led to Not Proven verdicts were also seen as being more difficult in comparison to decisions that led to other verdicts. Smithson et al.'s [7] research, therefore, suggests that the introduction of a third verdict has an impact on the verdict most likely to be

chosen, the difficulty of the task, and does not deter jurors away from Guilty verdicts. Their research, however, lacked a theoretical underpinning and used an Australian sample who may not be accustomed to the Not Proven verdict.

This study considers the impact that a three-verdict system may have on jurors, but has not yet examined whether or not a three-verdict system would be beneficial. Jackson [2] used discursive analysis and semiotic analysis (the study of symbols) to show that the aims of a juror may differ in comparison to the legal system, and that the inclusion of the Not Proven verdict may change the meaning of the Not Guilty verdict. First, the layperson may see the jury as ontological fact finders, whereas legal professionals may see jurors as cognitive processors who use information to prove a certain verdict [2]. Therefore, the Not Guilty verdict may mean one thing for a lawyer (that proof was lacking) and another for a juror (that the truth surrounding the innocence of the defendant has been discovered).

In addition, Jackson [2] suggests that the definition of the Anglo-American verdicts may differ when another verdict is available. When a Guilty verdict is given in a two-verdict system, it could be said that the information supplied allowed guilt to be proven beyond reasonable doubt[2]. Conversely, Not Guilty verdicts in a two-verdict system can be given either when a jury believes that the person was truly innocent, or when a jury believes that the defendant was truly Guilty, but the evidence was lacking in relation to a conviction [2]. This would then mean that acquitted individuals who are truly innocent might face social sanctions as the public may think that they were actually Guilty [2].

One way to counter the above criticism of Not Guilty verdicts in a two-verdict system is to introduce the Not Proven verdict. The introduction of the Not Proven verdict would mean that individuals who have been given a Not Guilty verdict would face no social and no legal sanctions, as a jury of their peers have shown that they thought that the defendant's innocence

has been proven [2]. Therefore, the definition of the Not Guilty verdict is much more potent in the Scottish three-verdict system because juries have to, consciously, bypass both a Not Proven verdict to give it [2]. Individuals who are given the Not Proven verdict will be free from legal sanctions as they are acquitted, yet may still face social sanctions due to the lack of confidence from the jury in relation to their innocence [2, 5]. Hope et al.[5] highlighted that jurors that are given the Not Proven verdict may face social sanctions, as 92% of their participants believed that defendant that were given the Not Proven verdict may suffer from stigma. In other words, the introduction of the Not Proven verdict may change what jurors interpret the Not Guilty verdict to mean. This may then mean that in a three-verdict system that the Not Guilty verdict is not an option given if the evidence does not show proof of guilt beyond reasonable doubt, and is rather a verdict given if the juror believes the evidence has shown proof of innocence. Further, the Not Proven verdict may replace the Not Guilty verdict and thus become the verdict that is given when a failure to meet the burden of proof has occurred. This may not have been the legal intention of the Not Proven verdict, but it may be the reality of what is occurring in the mind of a juror. The definition of the Guilty verdict should not differ between either of the verdict systems, however, as the Not Proven verdict does not directly compete with said verdict [2]. This is because both the Not Proven verdict and the Guilty verdict have differing legal outcomes and therefore jurors are likely to see them as opposing verdicts [2].

The inclusion of the Not Proven verdict arguably makes the Scottish system a more common sense option in comparison to its Anglo-American counterpart for assessing the guilt of a defendant [1]. This is in the sense that it allows jurors to communicate to society what they believe the evidence has provided proof of (i.e., innocence, guilt, or nothing) rather than forcing a binary choice of guilty versus not when the juror may not feel that the evidence is sufficient to make that binary choice. One problem with the Not Proven verdict, however, is

that the public may wrongly think that an innocent defendant who has been given the Not Proven verdict is Guilty [2]; thus, the Not Proven verdict does not solve the interpretation problems of the Not Guilty verdict (as mentioned earlier), it simply diverts the issue. In addition, Scottish jurors may confuse Not Proven verdicts with Not Guilty verdicts because of the similarities in legal outcomes, and this may cause some innocent individuals to suffer unwarranted social sanction [2]. In other words, the Not Proven verdict may allow defendant to fairly be acquitted, but its ambiguity may cause defendants who receive said verdict to face social sanctions from the public[2]. Another criticism of the not proven verdict relates to the fact that it may undermine the presumption of innocence, thus meaning that it might break article 6 of the European Convention on Human Rights [8].

Bhatia [6], Hope et al. [5] and Smithson et al. [7] suggest that the introduction of an additional, middle option has an impact on which verdict/option is chosen. These results taken together suggest that jurors, and general decision makers, may be violating the regularity principle. The regularity principle states the following: "the addition of an option to a choice set should never increase the probability of selecting an option from the original set" [9]. The reason then that the regularity principle may have been violated was because if a Not Guilty verdict can be given in an initial choice set, then the same verdict should be given when the same (or similar) evidence is presented, regardless of the number of verdicts available [5]. In addition, Hope et al.'s [5] study showed that the perception of guilt surrounding a defendant does not change across the verdict systems, which also highlights that jurors may be deviating from rational/normative models of decision making[10]. This is because normative/rational models would predict that a similar amount of information integration should lead to the same verdict being chosen [11], whereas Hope et al.'s [5] study found that the verdict system influenced the verdict chosen but not the perception of guilt surrounding the defendant given.

In summary, the current section has highlighted that the introduction of an additional option, such as the Not Proven verdict, can have an impact on verdict choice [6, 5, 2]. In addition, Hope et al. [3] suggested that jurors violate the regularity principle, as the likelihood of choosing a Not Guilty verdict in an initial choice set (i.e., the two-verdict system) should not be altered by the inclusion of the Not Proven verdict. Hope et al. [5] also found that estimates of guilt surrounding the defendant did not differ significantly across the two-verdict systems, which further highlights that jurors may be deviating from rational norms. Therefore, the current research will test whether or not jurors deviate from rational norms by comparing belief of guilt ratings and the verdicts given across the Anglo-American system and the Scottish system. The current study has one research aim, which is to test whether or not any differences in relation to the total belief of guilt score exist between the different verdict systems. This research aim was developed to test if jurors are deviating from rational norms.

The hypothesis for this section was:

• Significantly fewer Not Guilty verdicts will be given in the three-verdict condition in comparison to the two-verdict condition.

Method

Design.

This quasi-experiment adopted a one way within subjects design, the number of verdicts available (two vs three-verdict system) was used as a factor. The first dependent measure used in the current study was the total belief of guilt score and the second dependent measure used was the verdict given variable, which varied depending on the number of verdicts available condition.

Counterbalancing of vignettes and factors.

Participants heard two vignettes in the current investigation; these vignettes were counterbalanced across participants. the number of verdicts available factor was also counterbalanced, half of the participants could give one of three-verdicts in the first vignette they heard (Guilty, Not Guilty or Not Proven) and one of two-verdicts in the second vignette they heard (Guilty or Not Guilty), and the other half of participants could one of two-verdicts in the first vignette they heard and one of three-verdicts in the second vignette they heard; the number of verdicts available condition that participants were placed in first was randomly assigned.

Participant.

The current study included 128 participants (female = 98). Participants were aged between 18 and 61 years old (M =24.93, SD = 8.02). In this sample, 114 of the participants were students, the sample also consisted of other occupations, including but not limited to: academics, semiskilled workers and the unemployed. Only two inclusion/exclusion criteria were applied:

• Are you eligible to vote?

• And, are you eligible to be on a jury?

Participants were recruited via opportunistic sampling through placing advertisements on a number of websites (e.g., Gumtree, Call for Participants, and Craigslist, Facebook and Twitter).

Measure of guilt.

Guilt was measured by asking participants to mark their belief of guilt on a visual guilt rating scale; see Figure 1.

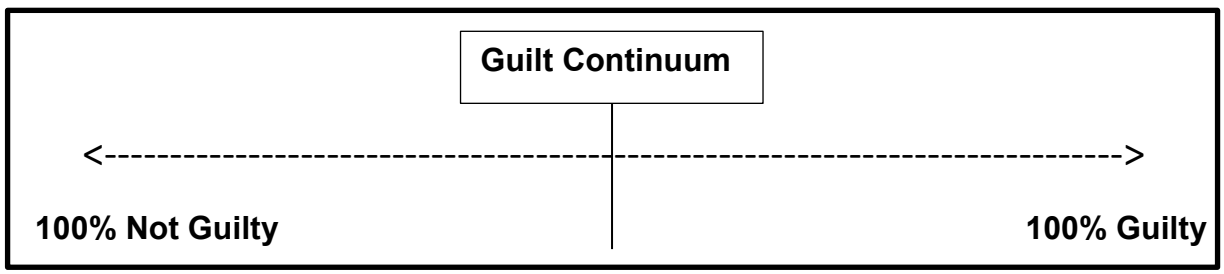

Figure 1. Guilt rating scale.

The guilt rating scale, which is a visual analogue scale, is 16 cm long; the bi-directional arrows within it are 14 cm long (each arrow spans 7cm from the centre point to the end). The far left of the guilt rating scale (marked with this arrow <) symbolises 100% Not Guilty and the far right of the guilt rating scale represents (marked with this arrow >) 100% Guilty. This scale was used so that participants could visually show how guilty they thought the defendant was, and so that the researcher could code these visual ratings of guilt into ratio data by measuring, with a ruler, how far the participants had marked their belief of guilt ratings from the Not Guilty arrow point on the far left. For example, if a participant marked their belief of guilt on the tip of the not guilty arrow, then their belief of guilt score would be zero, whereas if they marked said belief on the tip of the Guilty arrow, then their belief of guilt score would be 14. The rating of guilt score therefore had a possible range of 0-14 cm. Participants rated their belief of guilt after the opening statement, after each piece of evidence, and after each closing statement. All of these respective belief of guilt ratings were then summed to create a total belief of guilt score; this allowed the perception of guilt throughout the whole vignette to be measured. This total belief of guilt score had a potential range of 0-154; the higher this score was, the higher the belief of guilt was. Each participant gave two total belief of guilt scores, one for each of the two vignettes the participants heard.

Vignettes.

Development of vignettes.

Two vignettes were developed, each with two versions to allow counterbalancing of information. The vignettes utilised in the current study were fictional, but were inspired from real life trials, and the information from these real life trials was gathered from court transcripts and newspapers. The researcher also attended a real life high court trial to make

sure that the vignettes used were structured like a real life trial. Literature on vignette development was consulted when designing the vignettes [11, 12]. The two vignettes were designed to be the same length (i.e., 962 words long). Each of the vignettes were designed to contain an opening statement (130 words long for each of the vignettes) that provided context, four pieces of prosecution evidence, four pieces of defence evidence, and two closing statements (one from each stance).. The information presented in the two vignettes were similar in narrative: a female victim, a male defendant, a homicide trial, similar motives and the same evidence types were heard (e.g., DNA evidence, eyewitness testimony, expert testimony and secondary confessions). The only information that differed across the vignettes were names, ages and the specific details of the evidence and opening statements to improve generalizability. The ages of the victims and defendants, however, were comparable across the vignettes.

Audio vignettes.

Audio vignettes were constructed by employing a voice actor to read out the written vignettes in a quiet room at Edinburgh Napier University. Vignette one was either 5 minutes and 49 seconds long or 5 minutes and 51 seconds long, depending on the counterbalancing version of the vignette. Vignette two was either 5 minutes and 56 seconds long or 6 minutes and 2 seconds long, depending on the counterbalancing version of the vignette. Regardless of the slight time differences from editing, each version of a vignette provided the same pieces of information, and only the order of the evidence varied. The vignettes were recorded using theaudio recording app *Dictaphone – Audio Recorder* on an iPhone 5. The audio recording was then edited using the software *Audacity*.

Audio playback device.

The participants heard the vignettes in a quiet Psychology laboratory at Edinburgh Napier

University. The audio vignettes were saved as an mp3 file, and played using the software *Windows Media Player*. The participants heard the vignettes through *Labtec Spin 95* speakers.

Verdict judgments.

Once all the evidence had been provided in a vignette, participants were asked to give a final verdict, the verdicts that were available to them differed on the condition, thus allowing the within-subjects factor of the number of verdicts available (two-verdict system vs. three verdict system) to be created. Over this quasi-experiment, participants were presented with one vignette where they could give one of three-verdicts and another vignette where they could give one of two-verdicts. Counterbalancing of vignettes and factors' for more information on how the verdicts available factor was counterbalanced across participants. Participants were told which verdict options were available to them before starting each of the vignettes.

Information sheet, consent form and debrief sheet.

Standardised information sheets and debrief sheets made participants aware of the exclusion/ inclusion criteria, ethical issues (such as the right to withdraw and confidentiality), and who to contact if they need additional support. In addition, the consent forms allowed participants to consent to partaking in the pilot.

Procedure.

Participants read the standardised information sheet and completed the consent form within a quiet room in the Psychology Laboratory at Edinburgh Napier University. Once the consent form was signed, participant filled out the demographics questionnaire. Participants were then played the opening statement for the first vignette and were asked to state an initial/prior

rating of guilt by marking on the guilt rating scale (see Figure.1). They then heard the first piece of prosecution evidence and were asked to mark their belief of guilt rating again on another, fresh copy of the guilt rating scale. After this, participants heard the first piece of defence evidence, they were then asked to mark their belief of guilt on the guilt scale rating. This procedure of hearing evidence and giving a belief of guilt rating on the guilt rating scale was continued for the remaining six pieces of evidence (three for the prosecution and three for the defence). Once all of the evidence had been heard, the participants listened to two closing statements: one from the prosecution, which was heard first, and one from the defence. After hearing each of the closing statements, participants were asked to give their belief of guilt on the guilt rating scale. The participants were then given the following instructions from the judge: "Remember, to give a Guilty verdict you must think that the defendant was guilty beyond reasonable doubt". Finally, participants were asked to give a verdict, the number of verdicts available to them depended on the condition (two-verdict condition = Guilty or Not Guilty; three-verdict condition = Guilty, Not Guilty, or Not Proven). Once the first vignette was completed, participants repeated this procedure with a second vignette. Once the participants had finished, they were given a debrief sheet and were asked to confirm that they were happy for their data to be analysed by ticking a box; no participants chose this withdrawal option.

Ethics.

The current study was granted ethical approval by the Research Integrity Committee in the School of Applied Sciences at Edinburgh Napier University.

Results

First, ANOVA will be reported which tested the effects that the number of verdicts avaialible had on the total belief of guilt score. Finally, two separate chi squares which investigated the

relationship that the number of verdicts available had with the verdicts given variable will be presented.

Data Treatment for the total belief of guilt score.

For each of the vignettes, all of the belief of guilt scores given were summed to create the total belief of guilt score. This meant that for the total belief of guilt score analysis each participant was counted twice.

Descriptive statistics for the total belief of guilt score across the number of verdicts available.

Before inferential statistics could be conducted, descriptive statistics were explored. The data were found to be normally distributed, with minimal outliers present, thus allowing parametric testing to be applied. Table 1 presents the descriptive statistics for the number of verdicts available (two versus three).

Table 1

Descriptive statistics for the number of verdicts available factor across the total belief of guilt score.

	Total belief of guilt score	
Variable Name	Mean	Standard Deviation
The Number of Verdicts Available Condition:		
Two	82.90	24.92
Three	80.75	25.36

Table 1 highlighted that responses in the two-verdict condition were slightly higher in comparison to the three-verdict condition.

Categorical data treatment.

Initially, the number of verdicts available factor was treated as a within-subjects factors. However, this design violated the assumptions of the Chi-Square test of independence. The current researcher, therefore, replicated the analysis of Smithson et al. [7], who also investigated the effects of the inclusion of the Not Proven verdict, by only analysing the first verdict returns. These first verdict returns were counterbalanced across the number of verdicts available factor. In addition, the first verdict returns were also counterbalanced across the different vignettes. This data treatment allowed the number of verdicts available factor to be treated as a between-subjects factors, thus allowing the assumptions of the Chi-Square test of independence to be met.

Testing the effects that each factor the number of verdicts available had on the total belief of guilt score.

Results comparing the total belief of guilt score across the factors.

To test the effects that the number of verdicts available factor had on the total belief of guilt scale, one paired samples t-test was conducted. The paired samples t-test investigated the effects of the number of verdicts available factor on the total belief of guilt score, and it was found that no significant main effect existed [t (127) = .756, p = .45, d = .12].

Investigation of the association between the verdict given variable and the number of verdicts available.

When investigating the relationship between the verdict given and the number of verdicts available, two separate Chi-Square tests of independence were conducted. In the first, Not Proven verdicts were combined with Not Guilty verdicts. In the second, Not Proven verdicts were combined with Guilty verdicts. This was conducted as Not Proven verdicts could not be given in the two-verdict system, yet the Chi-Square test of independence would have still produced an estimated frequency for Not Proven verdicts within the two-verdict system, this would have nullified the results of said test. In addition, Hope et al. [5] used a similar method of analysis when testing the relationship between the verdict that was given and the number of verdicts available.

When Not Proven verdicts were combined with Guilty verdicts, it was found that jurors in the three-verdict condition (23.5%) were significantly less likely to reach a Not Guilty verdict in comparison to jurors in the two-verdict condition (76.5%) [$x2$ (1) = 23.76, p <.001, ϕ =.43]. In addition, when Not Proven verdicts were combined with Not Guilty verdicts, it was found that jurors in the three-verdict condition (41.9%) gave fewer Guilty verdicts than jurors in the

two-verdict condition (58.1%), this association was found not to be significant though [$x2$ (1) = 1.72, p =.19, ϕ =.12]. In summary, the inclusion of the Not Proven verdict reduces the likelihood of the Not Guilty verdict being chosen.

Discussion

Statement of principal findings

The current section will first discuss whether or not the number of verdicts available influenced the number of Not Guilty verdicts given. Then, this section will discuss if the number of verdicts available influenced juror perceptions of guilt. Finally, the implications of these results will be discussed.

First, it was found that jurors in the three-verdict condition were significantly less likely to give Not Guilty verdicts in comparison to jurors in the two-verdict condition, thus allowing the hypothesis to be accepted. Both Hope et al. [5] and Smithson et al. [7] found the same result in their own previous research. In addition, it was found in the current study that Guilty verdicts were equally likely across the two conditions. These two results taken together support Smithson et al.'s [7] conclusion that the Not Proven verdict does not sway jurors away from the Guilty verdict, and rather that the Not Proven verdict decreases the number of Not Guilty verdicts given. This decrease in Not Guilty verdicts when the Not Proven verdict is available highlights that jurors are deviating from the regularity principle, which is a key component of many rational choice models [5]. This is because if jurors are shown the same/similar pieces of evidence, it would be expected that the initial set of verdict options (i.e., Guilty and Not Guilty) should be given with the same frequency regardless of whether a Not Proven verdict is included or not. The current section will now discuss possible explanations for this deviation from what the rational choice model would expect.

In addition, the current study wanted to investigate if the different verdict systems had a significant effect on the perception of guilt; no significant difference was found. Hope et al. [5] found that juror estimates of guilt surrounding the defendant did not differ in the two-verdict condition in comparison to the three-verdict condition, thus the current study's findings provide support to Hope et al.'s [5] results.

Meaning of the study: possible mechanisms and implications for clinicians or policymakers

The results from the current investigation suggested that asymmetrical dominance (see Bhatia [6]) did not occur in the current study, as the availability of the similar, yet recessive, Not Proven verdict did not increase the number of Not Guilty verdicts given. The decrease in the number of Not Guilty verdicts given in the three-verdict system in comparison to the two-verdict system may indicate that the Not Guilty verdict is interpreted differently in the three-verdict system than it is the two-verdict system [2]. The Not Guilty verdict may be given on two scenarios in the two-verdict system: 1) if the juror truly believes the defendant is innocent; and 2) when the juror believes that the defendant is probably guilty, but there is not enough evidence to convict [2]. However, in the three-verdict system, jurors may only give a Not Guilty verdict when the juror truly believes that the defendant is innocent, and any doubt concerning the innocence of the defendant may lead to a Not Proven verdict [5], [2]. Further, the inclusion of the Not Proven verdict may allow members of the jury to communicate their belief of guilt more adequately to the courtroom [2].

This change in interpretation of the Not Guilty verdict may provide support for bounded rationality. Simon [13] coined the term bounded rationality, and suggested that both the environment and the decision maker's cognition interact when a decision is being made. Therefore, a change in the number of verdict options available may change the juror decision

making context, which may change how a juror interprets the Not Guilty verdict (i.e., a change in cognition) and this may then influence the final verdict that is reached. The Guilty verdict frequency did not significantly differ across the different conditions, however, which might suggest that the introduction of the Not Proven verdict does not change how the Guilty verdict is interpreted [2]. This may be because the Not Proven verdict is an additional acquittal verdict, and has no direct competition with Guilty verdicts [5], [2]. One negative implication from the findings in the current study is that truly innocent individuals in the three-verdict system may have a shadow placed over their innocence [2], as they will be less likely to receive a Not Guilty verdict in comparison to their two-verdict counterparts. Further, Hope et al. [5] found that their participants believed that a defendant that was given a Not Proven verdict would face more social stigma than a defendant that was given a Not Guilty verdict, and this could have severe implications for truly innocent defendants given Not Proven verdicts in serious crimes such as homicide. In addition, the inclusion of the Not Proven verdict may mean that jurors do not start a trial with a "presumption of innocence", which may limit how fair a trial the defendant is given, and this may break article 6 of the European Convention on Human Rights [8, 5]. Therefore, the Not Proven verdict may help jurors to communicate their belief of guilt to the court, but said verdict may have a negative impact on both human rights and truly innocent individuals.

In summary, the current study showed that the number of verdicts available did not have a significant impact on juror perceptions of guilt, but did influence the amount of Not Guilty verdicts given. These results have two theoretical implications. The first was that jurors do not make rational decisions. Jurors were presented with the same/similar pieces of information, and the perception of guilt did not differ across the two different verdict systems. Therefore, if jurors were being rational, it would be expected that the initial decision set (Guilty and Not Guilty) would be chosen with the same frequency regardless of whether

or not the Not Proven verdict was included. This did not occur, however, suggesting that jurors were deviating from the regularity principle, which is part of many rational choice models [5]. The second was that bounded rationality might explain why jurors deviate from what would be expected in rational choice models. Simon [13] and, more recently, Gigerenzer and Goldstein [14] have suggested that cognitive processes are simplistic and have evolved to adapt to an ever changing environment. Therefore, if the juror environment changes (i.e., a Not Proven verdict is included) then cognition may change (i.e., the interpretation of the Not Guilty verdict may change), meaning that similar perceptions of guilt may lead to different verdicts being chosen (Not Proven over Not Guilty).

Strengths and weaknesses of the study

One limitation of the current research was related to ecological validity, as a number of factors that occur in a courtroom cannot be replicated in juror decision experiments. For instance, an experiment cannot last as long as a real trial does, participants volunteer to take part in juror studies whereas an authoritative institution requests real life jurors, and mock jurors make decisions in a laboratory at a university rather than through a deliberation in a courtroom. Therefore, the extent to which the current study mirrors a real life courtroom may be limited, and the researchers acknowledge this. In addition, the current study only focussed on juror decision making rather than focussing on jury decision making and the deliberation process. This once again limits the extent to which the jurors in the current study experienced a similar environment to real life jurors.

However, the current design and focus (i.e., on jurors rather than jury) was chosen for a number of reasons. First, jurors were chosen to be studied rather than juries as this was only the third paper investigating how jurors use the Not Proven verdict (and was the first experiment on said topic in the last 10 years). Therefore, the authors wanted to investigate

how individual jurors used the Not Proven verdict before they subsequently studied how the Not Proven verdict was used in deliberations.

Second, an experimental design was chosen as this allowed the researchers to manipulate how many verdicts were shown to the jurors, and was the only method open to the authors as real life trials cannot be studied. As previously mentioned these trials were shorter than real life trials and the trials were displayed to the participants through audio rather than the jurors seeing the trials. Nevertheless, a number of previous pieces of research have highlighted that the ecological validity of mock trials does not have an impact on the decision outcome or process [15]. For instance, research by Pesdek, Avila-Mora and Sperry [15] showed that culpability ratings were not impacted on by the medium (audio vs. visual) in which the cases were presented. Further, psychologists such as Watt and Quinn [16] have suggested that the environment does not have an impact on cognitive mechanisms, such as decision making, and rather the cognitive mechanism will act the same regardless of the environment. In other words, jurors will use the same decision processes in a mock setting as they will in a real life setting; only the experience will differ.

In addition to these theoretical justifications for the methodology, the authors designed the vignettes in line with best-practice methodological literature on designing vignettes for research, as describes in detail in the method section. Information was collected about trials from newspapers and court transcripts, this information was then used as inspiration to create the vignettes, and finally all of the team provided input in the final design of the vignettes. While this was appropriate for the purpose and scope of the current research, future research should hope to extend the ecological validity of the current study. The researchers of this paper are also excited to see whether the results of the Chalmers, Leverick and Munro study, which is currently ongoing and has achieved the unusual privilege of accessing real jurors to investigate jury decision making, will coincide with our own.

In summary, only through simulation can juror and jury decision making be researched. Researchers can only see what variables/factors have an impact on the decision making processes and outcomes of jurors through controlled (quasi) experiments. In addition, real life jurors cannot be observed in the UK because of the contempt of court rule. Furthermore, mock juror (or jury) simulations are the best psychologists can do just now, and that is why the current methodology was chosen.

Unanswered questions and future research

A future line of enquiry from the current study relates to how other additional verdicts may influence verdict choice. For instance, the current study found that the inclusion of the Not Proven verdict significantly decreases the number of Not Guilty verdicts given, and Smithson et al. [7] found that the option of Manslaughter decreased the chances of the defendant being found guilty of murder, and decreased the amount of Not guilty verdicts given. Future research could investigate how other additional verdicts, such as the Not Guilty but insane – a special verdict [7], influence verdict choice. In addition, other jurisdictions have different legal proceedings; in Spain, the judge presents a series of propositions and the jury decides proven or not proven to each of said propositions [17]. Therefore, future research could investigate whether or not differences in legal proceedings influence the decision making of jurors.

Conclusion

In conclusion, the current study showed that the number of verdicts available did not have a significant effect on the total belief of guilt score. However, it was found that the inclusion of the Not Proven verdict did make jurors less likely to give a Not Guilty verdict; this has positive implications concerning juror's abilities to communicate their belief of guilt to the court. Furthermore, in tradition with bounded rationality, the current study has provided insight into the effects that the legal environment may have on both juror perceptions of guilt

and the verdict outcomes reached by jurors. Based upon this I would recommend that the Scottish legal system keep the bastard verdict.

Reference List

1. Broadbridge, S. The “Not Proven” verdict in Scotland (UK Parliamentary Briefing Papers No. SN02710). London, England: UK Houses of Parliament. 2009.

2. Jackson BS. Truth or proof?: The criminal verdict. Revue internationale de semiotique juridique. 1998 Oct 1;11(3):227-273.

3. Barbato JM. Scotland's Bastard Verdict: Intermediacy and the unique three-verdict system. Ind. Int'l & Comp. L. Rev.. 2004;15: 543-583

4. Duff P. The Scottish criminal jury: A very peculiar institution. Law and Contemporary Problems. 1999 Apr 1;62(2):173-201.

5. Hope L, Greene E, Memon A, Gavisk M, Houston K. A third verdict option: Exploring the impact of the not proven verdict on mock juror decision making. Law and human behavior. 2008 Jun 1;32(3):241-252.

6. Bhatia S. Confirmatory search and asymmetric dominance. Journal of Behavioral Decision Making. 2014 Dec 1;27(5):468-476.

7. Smithson M, Deady S, Gracik L. Guilty, not guilty, or…? Multiple options in jury verdict choices. Journal of Behavioral Decision Making. 2007 Dec 1;20(5):481-98.

(Kahan, 2015).

8. Council of Europe. European Convention of Human Rights (Protocol 14). Strasberg,

Germany: Council of Europe. 2010.

9. Trimmer PC. Optimal behaviour can violate the principle of regularity. Proceedings of the Royal Society of London B: Biological Sciences. 2013 Jul 22;280(1763):20130858.

10. Kahan, D. Laws of cognition and the cognition of law. Cognition.2015: 135, 56-60.

11. Ashill NJ, Yavas U. Vignette development: an exposition and illustration. Innovative Marketing. 2006;2(1):28-36.

12. Heverly MA, Fitt DX, Newman FL. Constructing case vignettes for evaluating clinical judgment: An empirical model. Evaluation and Program Planning. 1984 Jan 1;7(1):45-55.

13. Simon HA. Rational choice and the structure of the environment. Psychological review. 1956 Mar;63(2): 1-11.

14. Gigerenzer G, Goldstein DG. Reasoning the fast and frugal way: models of bounded rationality. Psychological review. 1996 Oct;103(4):650-669.

15. Pezdek K, Avila-Mora E, Sperry K. Does trial presentation medium matter in jury simulation research? Evaluating the effectiveness of eyewitness expert testimony. Applied Cognitive Psychology. 2010 Jul;24(5):673-90.

16. Watt R, Quinn S. It depends what you do in the laboratory. British Journal of Psychology. 2008 Aug;99(3):351-4.

17. Thaman SC. Europe's new jury systems: The cases of Spain and Russia. Law and Contemporary Problems. 1999 Apr 1;62(2):233-59.