

In press: *Psychology, Public Policy and Law*

**Beyond a Reasonable Doubt:
Ordinary People's Views of an Extraordinary Concept**

Mandeep K. Dhami, Middlesex University, London

Samantha Lundrigan, Anglia Ruskin University

Send correspondence to:
Professor Mandeep K. Dhami
Department of Psychology
School of Science and Technology
Middlesex University, London
The Burroughs, Hendon, NW4 4BT, UK
E-mail: m.dhami@mdx.ac.uk

Acknowledgements

Thank you to Noor Al-Adnani, Nurcan Aliriza, Holly Leak, Roberta Mocanu and Kramele Fino for assistance with data collection. We also thank Abbe Lake and Tess Dieseth for assisting with qualitative data coding. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission.

Author Note

Earlier stages of the present research were presented at the 2024 Meeting of the Law & Society Association in Denver, USA, and the 2025 European Society for Empirical Legal Studies Conference, Toulouse, France.

Abstract

Beyond a reasonable doubt (BRD) is the standard of proof required for conviction in criminal trials in Anglo-American legal systems. We examined how lay people ($N = 220$) define BRD in qualitative terms, and quantitatively measured their perceptions of existing definitions of this standard of proof as well as the grounds they believe create ‘a reasonable doubt’. The qualitative data suggests that although individuals used relevant synonyms for BRD (i.e., degree of doubt present, certainty or probability required, amount and strength of proof provided, and alternative explanations beyond the defendant being guilty), the concept is vague in individuals’ minds, has variable meanings across people, and its interpretation can increase or reduce the threshold for conviction. There was no significant difference between the *sure* and *firmly convinced* definitions of BRD (used in England and Wales, and the US Federal courts, respectively) in terms of self-reported understanding, confidence in application, and judged likelihood of wrongful conviction or wrongful acquittal. However, for each of these two definitions, greater self-reported understanding was significantly positively associated with greater confidence in applying the standard of proof, and both self-reported understanding and confidence were significantly negatively associated with judged likelihood of wrongful conviction and wrongful acquittal. Finally, a weak prosecution case and strong defence case were each rated as creating significantly greater grounds for a reasonable doubt than either victim and offender behavior or the punitive impact of a guilty verdict. The present findings have implications for jury instructions.

Keywords: Beyond reasonable doubt, standard of proof, jury instructions, jury decision-making

Introduction

Court decisions are typically made under conditions of uncertainty, largely because the available information (evidence) on which to base a decision is partial and fallible. Therefore, complete or 100% certainty is not required to render a decision. Rather, the degree of certainty required must pass a given legal threshold or standard of proof. These thresholds differ depending on the type of law being applied in a case (i.e., civil v. criminal) as well as the sorts of cases being heard (in a civil context). In Anglo-American legal systems, *beyond a reasonable doubt* (BRD) specifies the degree of belief in, or probability of, guilt required to decide on conviction in criminal cases (see Clermont, 2013; Shapiro, 1991). It is said to be a principle of due process that aims to minimize the likelihood of wrongful convictions (see Arkes & Mellers, 2002; DeKay, 1996), and is considered to be the highest legal threshold (e.g., see McCauliff, 1982).

Indeed, legal thresholds arguably lie along a continuum of probabilities from 0-1 (or 0-100%). For instance, BRD is often theorized to be equivalent to a .9 probability (or 90% chance) of guilt (see Blackstone, 1765; Laudan, 2003; McCauliff, 1982; Newman, 1993; *United States v. Fatico*, 1978). However, standards of proof such as BRD are typically not expressed numerically, and opinions differ as to whether they should be quantified, as well as what that quantification ought to be, and on what basis it should be conceptualized (e.g., Contarino, 2021; Dhimi, 2008; Franklin, 2006; Kagehiro, 1990; Stoffelmayr & Diamond, 2000; Tillers & Gottfried, 2006; Pi et al., 2020). We argue that, in its verbal expression, BRD does not reflect common parlance (see also Heffer, 2006; Tiersma, 2001). This is concerning given that lay people often serve on juries in criminal trials (e.g., see Kutnjak Ivković & Hans, 2021).

Research shows that lay people lack an accurate understanding of judicial instructions more generally (e.g., Reifman et al., 1992). Consequently, BRD may be a vague concept in

individuals' minds, and a concept that differs in its meaning across individuals. The concept may also be interpreted in ways that either increase or reduce the threshold for conviction. Others have also expressed concern over the difficulty that people may have in understanding BRD (e.g., Clermont, 2013; Elwork et al., 1982; Greene, 2016; Heffer, 2006; Power, 1999; Tiersma, 2001; Young et al., 2001), and scholars themselves grapple with how best to define BRD (e.g., see Shapiro & Muth, 2020). Despite this, courts generally believe that lay people understand BRD (see Court of Appeal in England and Wales in *R v Majid* [2009]; United States Supreme Court in *Victor v Nebraska* [1994]; High Court of Australia in *Green v The Queen* [1997]; New Zealand Court of Appeal in *R v Wanhalla* [2007]). In the present paper, we explore lay people's understanding of BRD.

Past Research on Lay Interpretations of Beyond a Reasonable Doubt

Past research has largely examined lay understanding of BRD *quantitatively*. Studies have employed different methods for eliciting individuals' numeric interpretations of BRD, both within and outside the context of case facts. Specifically, researchers have typically asked individuals to provide point estimates of BRD on a 0-1 probability (or 0-100%) scale, either yielding one 'best' (or 'peak') estimate, or a range of estimates from lowest to highest, including the 'peak' (or 'best') estimate (Dhmi, 2008; see also Clermont, 2013). Crucially, the resulting quantitative interpretations are shown to be meaningfully associated with verdicts (e.g., Dhmi, 2008; Lundrigan et al., 2016; Horowitz & Kirkpatrick, 1996; Kagehiro, 1990; Kagehiro & Stanton, 1985; Koch & Devine, 1999; Park et al., 2016), indicating that these methods provide predictively valid measures of BRD.

Several findings have emerged from this literature, all of which support concerns raised about the potential misunderstanding of BRD (e.g., see Dhmi, 2008 and Hastie, 1993). First, there is imprecision (fuzziness or spread) with which individuals quantitatively interpret BRD (Dhmi, 2008; Lundrigan et al., 2016; 2018). Imprecision is sometimes

referred to as intra-individual variability. For example, Dhimi (2008) observed that within the context of a manslaughter case, the average spread of interpretations of BRD (calculated by subtracting an individual's lowest interpretation from their highest interpretation) among a jury-eligible Canadian student sample was 52%. This rose to 59% outside the context of a case. Thus, an individual would be willing to convict based on a higher or lower threshold on different occasions.

Second, there is wide inter-individual variability in quantitative interpretations of BRD (e.g., Aimone et al., 2022; Dhimi, 2008; Hastie et al., 1983; Horowitz & Kirkpatrick, 1996; Lundrigan et al., 2016; 2018; Martin & Schum, 1987; Mueller-Johnson et al., 2018; Nagel, 1979; Nagel & Neef, 1979; Park et al., 2016; Simon, 1970; Simon & Mahan, 1971). For instance, Mueller-Johnson et al. (2018) observed that the range of 'peak' interpretations of BRD was from 50% to 100% across their sample of the British jury-eligible public. Thus, juries may find it difficult to reach a unanimous verdict because individual jurors differ in the threshold they apply even though all members of the jury agree on the case facts.

Third, on average, individuals' quantitative interpretations of BRD typically move the threshold for conviction higher up or lower down the probability scale. In some studies, interpretations fall above a .9 or 90% threshold for conviction (e.g., Dhimi, 2008; Hastie et al., 1983; Lundrigan et al., 2016; 2018; Martin & Schum, 1987; Mueller-Johnson et al., 2018). In other studies, interpretations fall below this threshold (e.g., Aimone et al., 2022; Dhimi, 2008; Horowitz & Kirkpatrick, 1996; Lundrigan et al., 2016; 2018; Magnussen et al., 2014; Nagel, 1979; Nagel & Neef, 1979; Park et al., 2016; Simon, 1970; Simon & Mahan, 1971; White & Cicchini, 2019). Most recently, Mueller-Johnson et al. (2018) observed an average 'peak' interpretation of BRD as being 92.3%, whereas Aimone et al. (2022) observed what they call an average "guilt threshold" of 73.6% among a sample of US adults. Thus, regardless of opinions on whether BRD should be quantified, it is clear that interpretations of

BRD may serve to increase or reduce the threshold for conviction in ways that subsequently affect rates of wrongful conviction (and arguably wrongful acquittal).

Defining the Standard of Proof for Jurors

Given concerns about the potential for lay people to misunderstand BRD and the consequent negative implications, efforts have been made to define this standard of proof for jurors. Some jurisdictions have implemented standardized definitions or jury instructions (Heffer, 2006; Hemmens et al., 1997; Power, 1999). For example, in England and Wales (Heffer, 2006) as well as New Zealand (Young, 2003), the term *sure* is used to substitute for BRD. In the US Federal courts as well as some US State courts the *firmly convinced* instruction is used (Hemmens et al., 1997). In some jurisdictions, such as in some States in the US, there may be multiple, often competing, official definitions of BRD (Hemmens et al., 1997). Finally, in other jurisdictions such as in Australia, courts may define BRD as they wish, on a case-by-case basis (see *Green v The Queen* 1997). This effectively means that BRD may be understood and applied differently by jurors across jurisdictions.

The fact remains that in all jurisdictions, BRD is either defined verbally (qualitatively) or is substituted for another term or phrase (see Hemmens et al., 1997). In addition to sometimes being lengthy, these qualitative definitions may be semantically complex and have poor grammatical structure. Definitions also often include vague words, hedging and/or boosting words, as well as double- or multiple-negatives, and tautological expressions. The following existing definitions of BRD provide examples of vagueness, hedging, boosting, negations and tautology, respectively (boldface added): “doubt based on reason and **common sense**” (Oregon Judicial Department, n.d.); “a doubt as **might cause** a reasonable person to hesitate” (Judicial Council of California (n.d.); “actual and **substantial** doubt” (Nebraska Supreme Court, n.d.); “doubt must **not** influence you to return a verdict of **not** guilty” (Florida Supreme Court, n.d.); “**a doubt that is reasonable**” (Michigan Supreme

Court Committee on Model Criminal Jury Instructions, n.d.). It is claimed that such factors serve to confuse and cause misunderstanding of the BRD standard of proof, and that psycholinguistic improvements to the definition of BRD can aid comprehension (Charrow & Charrow, 1979; Dhimi et al., 2015; Dumas, 2000; Elwork et al., 1982; Heffer, 2006; Severance et al., 1984; Solan, 2006; Power, 1999; Tiersma, 2001).

Indeed, the linguistic context in which BRD is defined can also either help or hinder understanding of the concept. For instance, Dhimi et al. (2015) found that when the definition of BRD focused on the word ‘doubt’ (e.g., as in an instruction used in the US State of Pennsylvania), quantitative interpretations of this standard of proof were reduced to, on average, 79%. In addition, the average fuzziness or spread of interpretations in individual minds was 80% (calculated by subtracting an individual’s lowest interpretation from their highest interpretation), and the average variability in ‘peak’ interpretations across individuals was 25%. By contrast, when the definition of BRD focused on the word ‘proof’ (e.g., as in an instruction used in the US State of Ohio), the interpretation of BRD was increased to, on average, 95%, and the average intra-individual imprecision (fuzziness or) spread and inter-individual variability of interpretations was reduced to 59% and 11%, respectively.

Furthermore, while BRD itself communicates the probability of guilt required for conviction, definitions of BRD may often refer to other verbal probability expressions. For example, the Supreme Court of Canada in *Lifchus* (1997, italics added) states that “...Even if you believe the accused is *probably* guilty or *likely* guilty, that is not sufficient....On the other hand you must remember that it is *virtually impossible* to prove anything to an *absolute certainty* and the Crown is not required to do so.” There is ample psychological research to demonstrate that people do not understand nor interpret verbal probabilities as intended by the communicator, and that these terms may even convey unintentional messages such as

whether an outcome will occur or not occur, as well as implicitly recommending a specific decision option (for a review see Dhimi & Mandel, 2022).

Therefore, for a number of reasons, existing efforts to help jurors understand BRD, by providing them with qualitative definitions, may be ineffective in aiding comprehension. Studies comparing definitions of BRD against a condition where BRD is left undefined and/or against other definitions of this standard of proof appear to support this claim (e.g., Aimone et al., 2022; Dhimi, 2008; Dhimi et al., 2015; Horowitz & Kirkpatrick, 1996; Kagehiro, 1990; Kerr et al., 1976; Koch & Devine, 1999; Mueller-Johnson et al., 2018; Nagel, 1979; Nagel & Neef, 1979; Park et al., 2016). Some jury instructions, including *sure* and *firmly convinced*, may increase the threshold for conviction (e.g., Dhimi, 2008; Kerr et al., 1976; Koch & Devine, 1999; Mueller-Johnson et al., 2018; Montgomery, 1998; Zander, 2000, but see Horowitz & Kirkpatrick [1996] and Park et al. [2016] who observed a reduced threshold under the *firmly convinced* definition). However, other definitions reduce the threshold for conviction (e.g., Aimone et al., 2022; Dhimi, 2008; Horowitz & Kirkpatrick, 1996; Kagehiro, 1990; Kassir & Wrightsman, 1979; Kerr et al., 1976).

Additionally, the imprecision of BRD in the minds of individuals and the variability in interpretation of this standard of proof across individuals remains, despite different definitions of the concept (Dhimi, 2008; Dhimi et al., 2015; Mueller-Johnson et al., 2018; Park et al., 2016). For example, Mueller-Johnson et al. (2018) reported that the average spread (or imprecision) of individuals' interpretations of BRD was around 40% under both the *sure* and *firmly convinced* definitions, and the average standard deviation of 'peak' interpretations (representing inter-individual variability) was around 10%.

While research on the quantitative meanings that people assign to BRD (undefined or variously defined) is amassing, there remains relatively little research on the *qualitative* meanings that people give to this standard of proof. Recognizing how individuals

(mis)understand BRD qualitatively can further underscore the need to define this standard of proof and may offer variations of terminology that researchers could empirically test, and which courts could then use in an empirically informed way.

The Present Research

The main aim of the present research was to explore how lay people define BRD in qualitative terms. Specifically, after measuring the average length of their definitions of BRD (i.e., the number of words used), we thematically identified the synonyms that individuals use for this standard of proof.

We also aimed to build on past quantitative research by further examining individuals' perceptions of BRD. We did this in two ways. First, we measured individuals' perceptions of the *sure* and *firmly convinced* definitions of BRD that are currently used in England and Wales and the US federal courts, respectively. Specifically, we measured: (a) self-reported ease/difficulty of understanding the standard of proof, (b) confidence in applying the standard of proof, (c) judged likelihood of wrongful convictions using the standard of proof, and (d) judged likelihood of wrongful acquittals. Self-reported ease/difficulty in understanding BRD can reveal whether people believe they can meaningfully apply it as required by the law. Subjective understanding has also been shown to be associated with juror stress (Bornstein et al., 2005; Miller & Bornstein, 2004). Meanwhile, confidence in applying the standard can indicate if people feel able to exercise their legal responsibility when making consequential decisions (i.e., to convict or acquit). While past research has examined the self-reported ease/difficulty in understanding the *sure* and *firmly convinced* definitions as well as confidence in applying the standard of proof under these definitions (e.g., Horowitz & Kirkpatrick, 1996; Mueller-Johnson et al., 2018; Park et al., 2016), to-date, no-one has measured perceptions of how these definitions could affect wrongful convictions and wrongful acquittals. Such perceptions could arguably affect

application of these definitions of BRD given that individuals may wish to avoid errors in judgment (Arkes & Mellers, 2002). Indeed, Park et al. (2016) found that individuals averse to making an error in terms of false acquittals were more likely to reduce the threshold for conviction, including when it was defined as *firmly convinced*.

Second, we added to past quantitative research by measuring the factors that individuals believe create grounds for ‘a reasonable doubt’. There is a large literature demonstrating the influence of a host of legal and extra-legal factors on juror and jury decision-making (see Devine & Caughlin, 2014; Devine et al., 2001; Mitchell et al., 2005). By contrast, there is much less literature directly measuring the effect of various factors on interpretations of BRD. These few studies (which all employ quantitative measures) suggest that the standard of proof may vary according to offence-related factors including offence type or seriousness (e.g., de Keijser et al., 2014; Martin & Schum, 1987; but for an exception see Lundrigan et al., 2018), evidential strength (e.g., Horowitz & Kirkpatrick, 1996; White & Cicchini, 2019 but for an exception see Park et al., 2016), potential penalty (e.g., Martin & Schum, 1987), and defendant-related factors (e.g., Conklin, 2020). Thus, there remains a need to directly understand what sorts of factors people believe create grounds for ‘a reasonable doubt’. An understanding of this issue can help courts instruct juries appropriately by encouraging them to ignore specific extra-legal factors.

Method

Transparency and Openness

The study was not preregistered. The survey and data for both the quantitative and qualitative analyses as well as supplementary materials supporting Appendices A and B are available from https://osf.io/ycvpd/?view_only=3991a389cdf949d79802411777c22a22

Participants

Two hundred and twenty jury eligible members of the population in England and Wales (i.e., those aged 18-75) volunteered to participate in the present research. Of those who provided demographic information, 59.4% said they were female, and 70.8% described themselves as White (with the remainder describing themselves as a member of an ethnic minority group). The mean age of the sample was 42.53 ($SD = 17.42$). Sixty-three percent said they had a university-level education, and 69.2% reported being employed. Ten percent of the sample said they had previously served on a jury, and participants rated their chances of attending jury service, if called in the future, as on average 75.24% ($SD = 33.50$).

Survey

Participants completed a survey comprising five sections. The first instructed participants that “In a criminal trial, the defendant is presumed innocent unless the prosecution has proved the defendant’s guilt beyond a reasonable doubt. Our past research shows that the term ‘beyond a reasonable doubt’ is not very well understood by potential jurors. Therefore, we would like to develop a definition of this term so that everyone understands it in the same way when it is used in criminal trials.” Participants were then asked to define ‘beyond a reasonable doubt’ in their own words using the sentence stem “Beyond a reasonable doubt is...”.

The second and third sections of the survey together informed participants that “In some countries, the judge defines beyond a reasonable doubt for jurors.” Following this, the second section presented participants with the *sure* instruction: “Proof beyond reasonable doubt is proof that makes you sure of the defendant’s guilty”. The third section presented them with the *firmly convinced* instruction: “Proof beyond a reasonable doubt is proof that leaves you firmly convinced of the defendant’s guilt. There are very few things in this world that we know with absolute certainty, and in criminal cases the law does not require proof that overcomes every possible doubt. If based on your consideration of the evidence, you are

firmly convinced that the defendant is guilty of the crime charged, you must find him guilty. If on the other hand, you think there is a real possibility that he is not guilty, you must give him the benefit of the doubt, and find him not guilty.” Due to an administrative error, the order of the two instructions was not counter-balanced (in the Discussion section we note this had limited impact on the findings). In each of these two sections, participants responded to four items that measured their perceptions: (a) “How easy/difficult do you think this definition would be for a juror to understand?” (b) “How confident do you think a juror would be in applying this definition to a case?” (c) “To what extent could this definition lead the jury to wrongfully convict an innocent person?” (d) “To what extent could this definition lead the jury to wrongfully acquit a guilty person?” All responses were provided on 4-point rating scales labelled at each point. For items (a) and (b) the labels were: extremely [difficult/unconfident], somewhat [difficult/unconfident], somewhat [easy/confident], extremely [easy/confident]. For items (c) and (d) the labels were: extremely likely, likely, unlikely, not at all likely. The responses for these latter two items were reverse coded for data analyses to ease interpretation of results.

In the fourth section of the survey, participants read: “There are many grounds for why a jury may have a reasonable doubt about the defendant’s guilt. How reasonable do you think each of the following doubts are?”. They were then with a sentence stem introducing 12 items (potential grounds): “Juries may have a reasonable doubt about the defendant’s guilt because...”. The 12 items were: (a) “The prosecution case against the defendant is weak.” (b) “The defendant does not behave like a criminal.” (c) “One or more witnesses seem to corroborate (support) the prosecution arguments.” (d) “There is not enough prosecution evidence against the defendant.” (e) “Finding the defendant guilty may lead to imprisonment.” (f) “The witnesses do not seem credible.” (g) “The defence case for the defendant is strong.” (h) “There is a lot of evidence for the defendant.” (i) “The victim does

not behave like he/she has been victimized.” (j) “There is no forensic evidence against the defendant.” (k) “The defence team may strongly contest (argue against) the prosecution team.” (l) “One or more witnesses seem to corroborate (support) the defence arguments.” Responses to each item were provided on 4-point rating scales labelled at each point from completely unreasonable, fairly unreasonable, fairly reasonable, to completely reasonable. Note that these items were chosen based on an a priori assumption that they could later be combined into four distinct categories as follows: ‘strong defence case’, ‘weak prosecution case’, ‘victim & offender behavior’, and ‘punitive impact’ of a guilty verdict.

The final section asked participants to provide the aforementioned demographic information, including gender, ethnicity, age, education level and employment status. In addition, participants indicated whether they had previously served on a jury, and their likelihood of serving on a jury in the future if called for duty.

Procedure

Participants were recruited from a variety of social clubs and social media sites through word-of-mouth as well as poster and social media advertising posts. Data was collected in three waves, with the first two being online and the last being via paper-pencil.¹ Ethical approval was obtained from the Department of Psychology Research Ethics Committee at first author’s University (application number: 3958). Participation was voluntary, anonymous, and followed informed consent procedures. Participants were debriefed after they completed the survey.

Analysis and Findings

Below, we report the analyses and findings of the quantitative data² followed by the qualitative data.

¹ In the online version, participants could not return to questions they had already completed, and in the paper version no-one appeared to change their answers.

² The quantitative data was analysed using IBM SPSS Statistics Version 28.0.0.0 (190).

Perceptions of Existing Definitions of Beyond a Reasonable Doubt

Table 1 presents the means and standard deviations of participants' perceptions of the *sure* and *firmly convinced* definitions of BRD in terms of self-reported ease/difficulty of understanding, confidence in application, judged likelihood of wrongful conviction, and judged likelihood of wrongful acquittal. Ratings of the two definitions were first compared using paired samples *t*-tests, with Bonferroni corrections applied to the alpha level for multiple testing ($\alpha = .01$). This indicated that, on average, there was no statistically significant difference in participants' perceptions of the *sure* and *firmly convinced* definitions, all $ps > .01$ (see Table 1).

Table 1. Means and Standard Deviations of Individuals' Perceptions of the *Sure* and *Firmly Convinced* Definitions of Beyond a Reasonable Doubt ($N = 209$).

	Sure		Firmly convinced		<i>t</i> -test
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Ease of understanding	2.79	.88	2.79	.84	$t(208) = .07, p = .941, d = .01$
Confidence in application	2.86	.77	2.85	.80	$t(209) = .15, p = .879, d = .01$
Wrongful conviction	2.41	.70	2.30	.73	$t(208) = 2.22, p = .027, d = .15$
Wrongful acquittal	2.42	.70	2.34	.71	$t(208) = 1.53, p = .127, d = .11$

Next, Pearson's correlations, with Bonferroni corrections applied to the alpha level for multiple testing (i.e., $\alpha = .005$), were used to reveal a significant positive association between greater perceived ease of understanding the definition and greater confidence in

applying the standard of proof for both the *sure* ($r = .62, p < .001$) and *firmly convinced* ($r = .73, p < .001$) definitions. In addition, for both definitions, self-reported understanding was significantly negatively associated with judged likelihood of wrongful conviction (*sure*: $r = -.49, p < .001$ and *firmly convinced*: $r = -.62, p < .001$) and judged likelihood of wrongful acquittal (*sure*: $r = -.51, p < .001$ and *firmly convinced*: $r = -.53, p < .001$). Similarly, confidence was significantly negatively associated with both judged likelihood of wrongful conviction (*sure*: $r = -.46, p < .001$ and *firmly convinced*: $r = -.66, p < .001$) and judged likelihood of wrongful acquittal (*sure*: $r = -.50, p < .001$ and *firmly convinced*: $r = -.60, p < .001$).

Perceived Grounds for ‘A Reasonable Doubt’

Before analysis, the 12 items representing potential grounds for ‘a reasonable doubt’ were combined (i.e., responses were averaged) into four distinct categories based on an a priori assumption that the items belonged to these categories. Cronbach’s alpha was then used to calculate the internal consistency of each category (where values can range from 0 to 1, with 1 indicating perfect consistency). The first category was labelled ‘strong defence case’ and included three items (i.e., g, h, l; $\alpha = .70$).³ The second category was ‘weak prosecution case’ and included four items (i.e., a, d, f, j; $\alpha = .52$).⁴ The third category included items b and i, and was labelled ‘victim & offender behavior’ ($\alpha = .61$). The remaining item (i.e., e) alone referred to the ‘punitive impact’ of a guilty verdict on the defendant. Table 2 presents the means and standard deviations of these four categories.⁵

³ As shown in the Supplementary materials, the inclusion of item k (i.e., “The defence team may strongly contest/argue against the prosecution team”) reduced the scale reliability, and so it was excluded from further analyses. We thank a reviewer who stated that this exclusion is understandable since the item is about argumentation whereas the others in the category are about evidence.

⁴ As shown in the Supplementary materials, item c (i.e., “One or more witnesses seem to corroborate/support the prosecution arguments”), when reverse coded, reduced the scale reliability and so it was excluded from further analyses.

⁵ Given that Cronbach’s alpha values were not above .70 (which is deemed acceptable) for two categories and only reached .70 for one category, we recommend that readers also refer to the Supplementary materials which show how individual items are related to one another. Additionally, in Table 2 we present the means and standard deviations for each item to show that responses on them were not vastly different from the category.

Table 2. Means and Standard Deviations of Perceived Grounds for a Reasonable Doubt.

	<i>M</i>	<i>SD</i>
Strong defence case	3.14	.57
(g) the defence case for the defendant is strong	3.25	.72
(h) there is a lot of defence evidence for the defendant	3.14	.79
(l) one or more witnesses seem to corroborate (support) the defence arguments	3.04	.65
Weak prosecution case	3.11	.49
(a) the prosecution case against the defendant is weak	3.19	.76
(d) there is not enough prosecution evidence against the defendant	3.35	.69
(f) the witnesses do not seem credible	3.03	.75
(j) there is no forensic evidence against the defendant	2.88	.85
Victim & offender behaviour	2.00	.80
(b) the defendant does not behave like a criminal	2.00	.94
(i) the victim does not behave like he/she has been victimized	2.00	.94
Punitive impact	2.03	.97
(e) finding the defendant guilty may lead to imprisonment		

The categories were then compared using paired samples *t*-tests, with Bonferroni corrections applied to the alpha level for multiple testing (i.e., $\alpha = .008$). This revealed that, on average, a weak prosecution case was rated as a significantly greater ground for ‘a reasonable doubt’ than either victim and offender behavior ($t[200] = 17.40, p < .001, d = 1.23$) or the punitive impact of a guilty verdict on the defendant, $t(199) = 13.94, p < .001, d =$

0.99. Similarly, a strong defence case was rated as a significantly greater ground for ‘a reasonable doubt’ than either victim and offender behavior ($t[199] = 16.98, p < .001, d = 1.20$) or the punitive impact of a guilty verdict, $t(198) = 13.63, p < .001, d = 0.97$. There was no significant difference in ratings between a weak prosecution case and a strong defence case, and nor was there a significant difference between individual victim and offender behavior and punitive impact of a potential guilty verdict, $ps > .008$.

Lay Definitions of Beyond a Reasonable Doubt⁶

Participants’ qualitative definitions of BRD, which were provided in response to the open-ended question at the beginning of the survey, were on average 22.66 ($SD = 22.14$) words in length, with a range from 2 to 159.

These definitions were read by the first and second author in order to identify the categories of synonyms that were used for BRD. This process resulted in five distinct and meaningful categories, two of which included a sub-category: (1) Degree of *doubt* that may be present in reaching a verdict, (2) *Certainty* or probability required to convict, which contained a subcategory (2a) Lack of certainty of guilt – pointing to innocence, (3) *Amount of proof* or evidence required to determine guilt, which also contained a subcategory (3a) Lack of proof or evidence – pointing to innocence, (4) *Strength of proof* or evidence required to determine guilt, and (5) Availability of *alternative explanations* other than defendant is guilty.

Two trained individuals then independently coded/rated each participant’s response into the above categories.⁷ Inter-rater (coder) reliabilities were calculated using Cohen’s (Weighted) Kappa which ranges from -1 to 1, with 1 indicating perfect agreement. The resultant values were as follows: .94 for doubt, .89 for certainty, .83 for lack of certainty, .90

⁶ Three participants did not answer the question asking them to define BRD.

⁷ Both coders were experienced in coding qualitative data. Additional training involved reading about the topic of BRD and the category definitions, as well as practice coding a sample of 20 participants’ responses. This was followed by a discussion with the first author, and agreement among the two coders.

for amount of proof, 1.00 for lack of proof, .87 for strength of proof, and .94 for alternative explanations. The two coders then met with the first author to discuss any disagreements until consensus was achieved.⁸

Table 3 presents the relative occurrence of each of these (sub)categories across participants. The most used synonyms for BRD were the degree of *doubt* that may be present in reaching a verdict and the *certainty* or probability required to convict (including the subcategory referring to lack of certainty of guilt – pointing to innocence). Appendix A presents instances falling into the aforementioned (sub)categories of synonyms (see Supplementary materials which indicate where in each participant's response an instance of the synonym appeared).

Overall, 80.2% ($n = 174$) of participants' responses contained at least one category of synonym ($M = 1.07$, $SD = .77$, range 0 to 5). There was a significant positive association of $r = .56$ ($p < .001$) between the number of synonyms in a response and the number of words.

As can be seen in Appendix A, variability in understanding of BRD across individuals is demonstrated by the range of instances falling into each of the categories (including subcategories). In addition, differences in the threshold for conviction are apparent in each category of synonym. For example, in the 'certainty or probability required to convict' category, some individuals used verbal or numeric expressions at a high- or top-point along the probability (or percentage) scale (i.e., "99%" or "100%", "*certain*") while others hovered around the mid-point (e.g., "*in the balance of probabilities*"). Instances in the 'degree of doubt that may be present in reaching a verdict' category also ranged from the top- to a lower point on the scale (e.g., "*no doubt(s)*" to "*when guilt is proved further than doubting*"). In the 'amount of proof or evidence required to determine guilt' category, instances similarly

⁸ The (separate) coding and (joint) agreement work was divided into five sessions to fit the coders' other work commitments. On a few occasions, in the joint agreement session, the two coders revisited responses where they initially agreed, and after discussion, they agreed on a different code (the relevant database in the Supplementary materials indicates where this occurred).

ranged from “*all the evidence*” to “*enough proof or evidence*”. In the ‘strength of proof or evidence required to determine guilt’ category, instances ranged from “*cast iron evidence and witnesses*” to “*the information gathered is strong enough*”. Lastly, in the ‘availability of alternative explanations other than defendant is guilty’ category, instances ranged from “*no other explanation*” to “*surpassed a reasonable excuse as to why they have not committed the crime*”.

Table 3. Relative Occurrence of Synonyms for BRD Across Individuals (N = 217).

Category	% (n)
• Subcategory	
Degree of <i>doubt</i> that may be present in reaching a verdict	25.8 (56)
<i>Certainty</i> or probability required to convict	19.8 (43)
• Lack of certainty of guilt – pointing to innocence	4.1 (9)
<i>Amount</i> of <i>proof</i> or evidence required to determine guilt	22.6 (49)
• Lack of proof or evidence – pointing to innocence	6.0 (13)
<i>Strength</i> of <i>proof</i> or evidence required to determine guilt	16.1 (35)
Availability of <i>alternative explanations</i> other than defendant is guilty	15.2 (33)

Further/exploratory qualitative analyses. In addition to the aforementioned coding of the synonyms that lay people used to define BRD, which was the main aim of the qualitative research, some participants’ responses also revealed other aspects of their understanding of the standard of proof. For instance, although participants were not instructed to define BRD numerically, 7.4% ($n = 16$) did so (see Supplementary materials). Here, nine

participants referred to the threshold being 99 or 100 percent, while the remainder referred to a threshold ranging from 50 to 90 percent.

Some participants used tautological expressions and/or existing definitions. Specifically, after excluding participants' use of the sentence stem i.e., "beyond a reasonable doubt is...", we found that 23.5% ($n = 51$) of participants used the term *doubt/doubts*, and 14.3% ($n = 31$) used the term *reasonable*. Only 4.6% ($n = 10$) participants used the term *sure*, and four of these referred to being 100% sure while one referred to being 99% sure. No-one in the present research used the term *firmly convinced*.

Similarly, in their efforts to define BRD, some participants referred to other aspects of jury decision-making in criminal trials. In other words, the synonyms for BRD were sometimes contextualized, namely in themes referring to the: (1) presumption of innocence, (2) burden of proof, (3) features of the judgment process, and (4) characteristics of the decision-maker. We used the same approach as mentioned earlier for coding these categories (except there were no practice cases). Inter-rater reliabilities were calculated using Cohen's (Weighted) Kappa and the resultant values were as follows: 1.00 for presumption of innocence, .92 for burden of proof, .91 for characteristics of the decision-maker, and .84 for features of the judgment process. Appendix B presents instances in each of these themes across participants (see Supplementary materials for the full set of instances).

Overall, 41.0% ($n = 89$) of participants' responses contained at least one of the above categories. Features of the judgment process was the most common contextual theme for definitions of BRD, mentioned by 18.0% ($n = 39$) of participants. Here, participants, for instance, described how jurors use the evidence presented, how they reason or think when making the decision, and/or how "*convinced*" in the evidence or decision they may be (see Appendix B). The presumption of innocence (11.5%, $n = 25$) and the burden of proof (11.1%, $n = 24$) were virtually equally likely to be mentioned. While most participants who referred

to these two categories, overall, appeared to have an accurate understanding of these legal concepts, a few demonstrated misunderstanding, for example, by presuming the defendant was guilty and by placing the burden of proof on the defence. Finally, a handful of few participants mentioned characteristics of the decision-maker (5.1%, $n = 11$) describing the kind of person who will/can make the decision, for instance, as being “reasonable” (see Appendix B).

Discussion

Beyond a reasonable doubt (BRD) is the highest standard of proof. However, this legal threshold is not an everyday expression, and is open to misinterpretation. We extended past research on BRD by *qualitatively* exploring lay people’s understanding of this concept. We also contributed to quantitative research on BRD in two ways: First, we measured a wider range of perceptions of the *sure* and *firmly convinced* definitions of BRD currently used in England and Wales and the US federal courts, respectively. Second, we directly investigated the factors that individuals believe create grounds for ‘a reasonable doubt’. Below, we discuss the findings as well as their implications, and propose avenues for further research.

Perceptions and Understanding of Beyond a Reasonable Doubt

Lay perceptions of official definitions. Our quantitative analyses of individuals’ perceptions of the *sure* and *firmly convinced* definitions of BRD revealed no significant difference between them in terms of perceived ease/difficulty in understanding the definition and confidence in applying the standard of proof given the definition. This is consistent with past research, which counter-balanced the order of these two instructions (Mueller-Johnson et al., 2018). Additionally, in an issue which has not been previously addressed, we observed that greater perceived ease of understanding the *sure* and *firmly convinced* definitions was significantly positively associated with greater confidence in applying the standard of proof given each definition. Finally, in a further extension to previous work, we found no

significant difference between the *sure* and *firmly convinced* definitions in terms of judged likelihood of wrongful convictions and wrongful acquittals under them. However, for both definitions, we did observe that greater self-reported ease of understanding and greater confidence were both significantly negatively associated with the judged likelihood of wrongful conviction and judged likelihood of wrongful acquittal.

Although the aforementioned findings are correlational, and need to be re-examined using an experimental design and/or where the order of instructions and questions is counter-balanced, they do signify the potential importance of ensuring that jurors subjectively believe they understand what BRD means. This may affect self-confidence in the ability to apply BRD, especially in a way that reduces erroneous decisions. Indeed, Park et al. (2016) found that individuals reduced the threshold for conviction in an effort to avoid false acquittals. To-date, no-one has measured both quantitative interpretations of BRD and perceived error rates under both the *sure* and *firmly convinced* definitions – an important aim for further research, given the adverse implications for juror stress (Bornstein et al., 2005; Miller & Bornstein, 2004).

Lay definitions. On average, lay people's definitions of BRD were shorter in length than those used by some courts (e.g., Florida Supreme Court, n.d.; Judicial Council of California, n.d.; Oregon Judicial Department, n.d.). The number of categories of synonym referred to in participants' definitions of BRD increased with the length of the definition. This suggests that efforts to define BRD can introduce vagueness (or imprecision) when the definition refers to different categories of synonym. Indeed, courts across several jurisdictions caution against over-complicating or over-explaining the concept so as not to further confuse jurors (see the Court of Appeal in England and Wales in *R v Majid* [2009]; the High Court of Australia in *Green v The Queen* [1997]; the New Zealand Court of Appeal in *R v Wanhalla* [2007]; and the United States Supreme Court in *Victor v Nebraska* [1994]).

The benefits and drawbacks of different lengths of jury instructions on BRD remain to be empirically demonstrated. On the one hand, longer instructions provide an opportunity to explain the concept in more detail. On the other hand, longer instructions may serve to place an additional cognitive burden on jurors.

The vast majority of individuals in the present study defined BRD using verbal expressions rather than numerically. As Dhimi and Mandel (2022) point out when discussing the predominance of verbal over numeric probabilities in the communication of uncertainty generally, “people find it easier and more natural to use words than numbers.” However, their review of the uncertainty communication literature also finds that ease of communication is not necessarily predictive of unambiguously understanding the message.

We also found that around one sixth to a quarter of individuals used tautological expressions when defining BRD i.e., containing the terms *doubt* and/or *reasonable*. This potentially highlights the serious effort made by the majority of participants to define BRD in their own words, as well as the difficulty that others had in doing so. Some courts similarly struggle to define BRD in words other than ‘doubt’ and ‘reasonable’ (e.g., Michigan Supreme Court Committee on Model Criminal Jury Instructions, n.d.).

No-one spontaneously used *firmly convinced*, and less than five percent used *sure*. A handful of the few who used *sure* referred to being either 99% or 100% sure. Although this threshold for conviction may be too high, the finding is consistent with past research eliciting quantitative interpretations of the *sure* definition of BRD (Montgomery, 1998; Mueller-Johnson et al., 2018; see also Zander, 2000). Overall, our findings suggest a potential lack of familiarity with the *firmly convinced* and *sure* definitions of BRD, which may partly explain why past research has found these two definitions to be ineffective in aiding comprehension of the standard of proof (e.g., Horowitz & Kirkpatrick, 1996; Koch & Devine, 1999; Montgomery, 1998; Mueller-Johnson et al., 2018; Park et al., 2016; Zander, 2000).

Lay people's definitions of BRD were grouped into five categories of synonyms (excluding two subcategories) i.e., degree of doubt present, certainty or probability required, amount and strength of proof provided, and alternative explanations beyond the defendant being guilty. One could claim that these categories are relevant to the standard of proof, and so the sample as a whole, appears to have grasped the gist of the concept. However, one could also say that specific instances in each category of synonym reveal misunderstanding, and that the findings tend to concur with past quantitative research, which concludes that BRD differs in its meaning across individuals, and its interpretation can increase or reduce the threshold for conviction (e.g., Aimone et al., 2022; Dhimi, 2008; Mueller-Johnson et al., 2018).

Specifically, the wide range of instances falling into each category of synonym highlights the variability in understanding of BRD across individuals, which poses challenges to jury (group) decision-making. Second, the range of instances in each category of synonym demonstrate differences in the threshold for conviction being set. For example, instances in the certainty or probability required to convict category spanned half of the probability (or percentage) scale i.e., from the high- or top-point down to the mid-point. Consistent with this, the subcategory (i.e., lack of certainty of guilt – pointing to innocence) contained instances of a threshold that required certainty as well as instances that appeared to be below the mid-point of the probability scale. Overall, the threshold for conviction may be increased or reduced in ways that can have an adverse effect on the likelihood of wrongful acquittals and wrongful convictions, respectively.

Under the aforementioned circumstances, it is tempting to prescribe alternative definitions of BRD, particularly in the face of recent calls to redefine this standard of proof (e.g., see Keane & McKeown, 2019). We believe that alternatives ought to be informed by

empirical evidence of ‘what works’.⁹ For instance, the following instance in the ‘certainty or probability required to convict’ category of synonym could be empirically tested: *strongly certain that there is a high probability* or some variant of it such as *very confident that there is a high probability*. This is because it differentiates between the judged probability of guilt and the juror’s confidence in that judgment. Quantitatively, measurements should be taken of the degree of imprecision, inter-individual variability, and the threshold implied, as well as subjective understanding and confidence, and perceived likelihood of errors. Qualitatively, a paraphrase methodology could be employed (see Charrow & Charrow, 1979). Comparisons should be made against when BRD is left undefined as well as against existing definitions. The findings of such an endeavour would provide useful evidence to inform jury instructions on BRD.

Perceived Grounds for ‘A Reasonable Doubt’

When prompted to consider what comprises grounds for ‘a reasonable doubt’, we found that individuals rated both a weak prosecution case and a strong defence case as creating significantly greater grounds for a reasonable doubt than victim and offender behavior or the punitive impact of a guilty verdict. These findings are consistent with the few past studies that have measured the effect of evidential strength on interpretations of BRD (e.g., Horowitz & Kirkpatrick, 1996; White & Cicchini, 2019 but for an exception see Park et al., 2016), but do not concur with studies pointing to an effect of potential penalty (e.g., Martin & Schum, 1987) and defendant-related factors (e.g., Conklin, 2020) on BRD. As mentioned earlier, there is little research on the factors which directly influence interpretations of BRD. Further work should be directed at what sorts of factors people

⁹ In addition, it has been argued that jurors should be instructed on the method of reasoning to be used alongside application of the standard (e.g., Clermont, 2013; Cicchini, 2019; Ho, 2021; Picnali, 2015). We propose that an evidence-based approach should also be applied to such instructions in order to ensure that they have the desired effects.

believe create grounds for ‘a reasonable doubt’ so that courts can use the findings to instruct juries by, for example, encouraging them to ignore specific extra-legal factors.

(Mis)understanding Jury Decision-Making in Criminal Trials

Although our study was not designed to test lay people’s comprehension of aspects of criminal trials other than the legal threshold to be applied (i.e., BRD) we found that some individuals do not fully understand the adversarial nature of criminal trials in Anglo-American jurisdictions, nor the fact that decisions are made under conditions of uncertainty. This was demonstrated by the top-end of the range of instances in the categories of synonyms referring to the degree of doubt present, amount and strength of proof provided, and alternative explanations beyond the defendant being guilty. For example, “*no doubt(s)*” being present, requiring “*all the evidence*” and “*cast iron evidence and witnesses*” to prove guilt, and “*no other explanations*” being available other than the defendant being guilty. According to Heffer (2006), ‘no’ is the most common modifier used for the word doubt used in natural language, but this does not fully account for the fact that individuals appeared to negate or be unaware of the adversarial and uncertain nature of criminal trials.

In addition, we observed that the synonyms for BRD were sometimes contextualized in ways that provided insights into how individuals understand jury decision-making in criminal trials. Briefly, most of the individuals who referred to the presumption of innocence and/or burden of proof, overall, appeared to have an accurate understanding of these two legal concepts. Only a few demonstrated incorrect understanding i.e., believing the defendant is to be presumed guilty, and placing the burden of proof on the defence. In the courtroom, aspects such as the presumption of innocence and the burden of proof are explained to jurors via judicial instructions, and so may be understood by all. However, as Charrow and Charrow (1979) point out, legal language is typically linguistically complex and consequently difficult for lay people to comprehend (see also Severance et al., 1984 and Tiersma, 2001). Future

research using a combined quantitative and qualitative approach could therefore be directed at how well these other aspects of jury decision-making in criminal trials are understood, especially, as research suggests that they may affect how quickly the threshold for conviction is reached (for presumption of innocence see Scurich & John, 2017).

Several of those individuals who described features of the judgment process for determining a verdict referred to being “convinced”. Few made explicit mention of using a rational strategy (i.e., evidence weighting and integration) and/or not being biased. Finally, the few individuals who referred to characteristics of the decision-maker, emphasized that they are a “reasonable person”, with some individuals elaborating on this. Both of these categories are present in official definitions of BRD (see Hemmens et al., 1997). Future research could investigate how belief in one’s ability to apply a particular decision process and/or identification with specific characteristics of the decision-maker affect interpretation and application of the standard of proof.

Conclusion

The standard of proof represents a fundamental aspect of court decision-making since it represents the threshold for arriving at a given decision under conditions of uncertainty. In criminal trials, the decision threshold encapsulated by ‘beyond a reasonable doubt’ is set high. However, past quantitative research concludes that BRD is often misunderstood by lay people who comprise the juries tasked with deciding whether a defendant is guilty of a crime or not. Our qualitative research contributes a rich layer of empirical evidence that concurs with the past work and extends it in useful ways. Given that the legal system confers great responsibility on lay people to make decisions, the system ought to be clearly articulate what it asks of them so that they can competently and confidently accomplish their task. This is especially since their have potentially severe consequences for a defendant’s liberty, a victim’s sense of justice, public safety, and society’s confidence in the legal system itself.

References

- Aimone, J. A., Hudja, S. and Law, W., North, C. M., Ralston, J., & Rentschler, L. (2022). An experimental exploration of reasonable doubt. <http://doi.org/10.2139/ssrn.3923809>
- Arkes, H. R., & Mellers, B. A. (2002). Do juries meet our expectations? *Law and Human Behavior*, 26, 625-639. <https://doi.org/10.1023/a:1020929517312>
- Blackstone, W. (1765). *Commentaries on the laws of England*. Oxford: Clarendon Press. <https://doi.org/10.1093/oseo/instance.00248899>
- Bornstein, B. H., Miller, M. K., Nemeth, R. J., Page, G. L., & Musil, S. (2005). Juror reactions to jury duty: perceptions of the system and potential stressors. *Behavioral sciences & the law*, 23(3), 321-346. <https://doi.org/10.1002/bsl.635>
- Judicial Council of California (n.d.). *CALJIC No. 2.90: Reasonable doubt*. In *California jury instructions*. https://www.courts.ca.gov/partners/documents/CALJIC_2_90.pdf
- Charrow, R. E., & Charrow, V. (1979). Making legal language understandable: A psycholinguistic study of jury instructions. *Columbia Law Review*, 79, 1306-1374. <https://doi.org/10.2307/1121842>
- Clermont, K. M. (2013). *Standards of decision in law*. Durham, North Carolina: Carolina Academic Press.
- Cicchini, M. D. (2019). Reasonable doubt and relativity. *Washington & Lee Law Review*, 76, 1443.
- Conklin, M. (2020). Reasonable doubt ratcheting: How jurors adjust the standard of proof to reach a desired result. *Notre Dame Law Review*, 95, 281.
- Contarino, N. (2021). Reasonable doubt: in law the highest burden in fact barely burden. *George Washington Law Review*, 89(3), 738-767.

- de Keijser, J.W., de Lange, E.G.M., & van Wilsem, J.A. (2014). Wrongful convictions and the Blackstone ratio: An empirical analysis of public attitudes. *Punishment & Society*, 16, 32-49. <https://doi.org/10.1177/1462474513504800>
- DeKay, M. (1996). The difference between Blackstone-like error ratios and probabilistic standards of proof. *Law & Social Inquiry*, 21, 95-132. <https://doi.org/10.1111/j.1747-4469.1996.tb00013.x>
- 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, 109-134. <https://doi.org/10.1037/law0000006>
- Devine, D. J., Clayton, L. D., Dunford, B. B., Seying, R., & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy and Law*, 7, 622-727. <https://doi.org/10.1037//1076-8971.7.3.622>
- Dham, M. K., & Lundrigan, S. (2025). *Beyond a Reasonable Doubt Project*
https://osf.io/ycvpd/?view_only=3991a389cdf949d79802411777c22a22
- Dhami, M. K. (2008). On measuring quantitative interpretations of reasonable doubt. *Journal of Experimental Psychology: Applied*, 14, 353-363. <https://doi.org/10.1037/a0013344>
- Dhami, M. K., Lundrigan, S., & Mueller-Johnson, K. (2015). Instructions on reasonable doubt: Defining the standard of proof and the juror's task. *Psychology, Public Policy and Law*, 21, 169-178. <https://doi.org/10.1037/law0000038>
- Dhami, M. K., & Mandel, D. R. (2022). Communicating uncertainty using words and numbers. *Trends in Cognitive Sciences*, 26, 514-526.
<https://doi.org/10.1016/j.tics.2022.03.002>
- Dumas, B. K. (2000). Jury trials: Lay jurors, pattern jury instructions, and comprehension issues. *Tennessee Law Review*, 67, 701-742.

Elwork, A., Sales, B. D., & Alfini, J. J. (1982). *Making jury instructions understandable.*

Charlottesville, VA: Michie/Bobb-Merrill. [https://doi.org/10.1016/0047-2352\(83\)90035-1](https://doi.org/10.1016/0047-2352(83)90035-1)

Florida Supreme Court. (n.d.). *Standard Jury Instruction 3.7: Reasonable doubt.*

<https://www.smithandeulo.com/reasonable-doubt/smithandeulo.com>

Franklin, J. (2006). Case comment—*United States v. Copeland*, 369 F. Supp. 2d 275

(E.D.N.Y. 2005): Quantification of the ‘proof beyond reasonable doubt’ standard. *Law, Probability and Risk*, 5(2), 159-165, <https://doi.org/10.1093/lpr/mgl017>

Greene, B. (2016). Reasonable doubt: Is it defined by whatever is at the top of the Google Search page. *The John Marshall Law Review*, 50, 933.

Hastie, R. (Ed.). (1993). *Inside the juror: The psychology of juror decision-making.*

Cambridge University Press. <https://doi.org/10.1002/bdm.3960070407>

Hastie, R., Penrod, S. D., & Pennington, N. (1983). *Inside the jury.* Cambridge, MA: Harvard

University Press. <https://doi.org/10.4159/harvard.9780674865945>

Heffer, C. (2006). Beyond ‘reasonable doubt’: The criminal standard of proof instructions as communicative act. *The International Journal of Speech, Language and the Law*, 13,

159-188. <https://doi.org/10.1558/ijssl.2006.13.2.159>

Hemmens, C., Scarborough, K. E., & del Carmen, R. V. (1997). Grave doubts about

‘reasonable doubt’: Confusion in State and Federal courts. *Journal of Criminal Justice*, 25, 231-254. [https://doi.org/10.1016/s0047-2352\(97\)00008-1](https://doi.org/10.1016/s0047-2352(97)00008-1)

Ho, H. L. (2021). Justification, excuse, and proof beyond reasonable doubt. *Philosophical*

Issues, 31, 146-166. <https://doi.org/10.1111/phis.12200>

- Horowitz, I. A., & Kirkpatrick, L. C. (1996). A concept in search of a definition: the effects of reasonable doubt instructions on certainty of guilt standards and jury verdicts. *Law and Human Behavior*, 20, 655-670. <https://doi.org/10.1007/BF01499236>
- Kagehiro, D. K. (1990). Defining the standard of proof in jury instructions. *Psychological Science*, 1, 194-200. <https://doi.org/10.1111/j.1467-9280.1990.tb00197.x>
- Kagehiro, D. K., & Stanton, W. C. (1985). Legal vs. qualified definitions of standards of proof. *Law and Human Behavior*, 9, 159-178. <https://doi.org/10.1007/BF01067049>
- Kassin, S. L. & Wrightsman, L. (1983). The construction and validation of a juror bias scale. *Journal of Research in Personality*, 17, 423-442. [https://doi.org/10.1016/0092-6566\(83\)90070-3](https://doi.org/10.1016/0092-6566(83)90070-3)
- Keane, A., & McKeown, P. (2019). Time to abandon ‘Beyond Reasonable Doubt’ and ‘Sure’: The case for a new direction on the criminal standard and why and how it should be used. *Criminal Law Review*, 6, 505-527.
- Kerr, N. L., Atkin, R. S., Stasser, G., Meek, D., Holt, R.W., & Davis, J. H. (1976). Guilt beyond a reasonable doubt: effects of concept definition and assigned decision rule on the judgments of mock jurors. *Journal of Personality and Social Psychology*, 43, 282-394. <https://doi.org/10.1037/0022-3514.34.2.282>
- Koch, C. M., & Devine, D. J. (1999). Effects of reasonable doubt definition and inclusion of a lesser charge on jury verdicts. *Law and Human Behavior*, 23, 653-674. <https://doi.org/10.1023/A:1022389305876>
- Kutnjak Ivković, S. S., & Hans, V. P. (2021). A worldwide perspective on lay participation. In S. Kutnjak Ivković, S. S. Diamond, V. P. Hans, & N. Marder (Eds.), *Juries, lay judges, and mixed courts: A global perspective*. Cambridge University Press.
- Laudan, L. (2003). Is reasonable doubt reasonable? *Legal Theory*, 9, 295-331. <https://doi.org/10.1017/S1352325203000132>

- Lundrigan, S., Dhami, M. K., & Mueller-Johnson, K. (2016). Predicting verdicts using pretrial attitudes and standard of proof. *Legal and Criminological Psychology*, 21, 95-110. <https://doi.org/10.1111/lcrp.12043>
- Lundrigan, S., Dhami, M. K., & Mueller-Johnson, K. (2018). A re-examination of the acquittal biasing effect of offence seriousness. *Psychiatry, Psychology and Law*, 25, 769-778. <https://doi.org/10.1080/13218719.2018.1478334>
- Magnussen, S., Eilertsen, D. E., Teigen, K. H., & Wessel, E. (2014). The probability of guilt in criminal cases: Are people aware of being 'beyond reasonable doubt'? *Applied Cognitive Psychology*, 28(2), 196-203. <https://doi.org/10.1002/acp.2978>
- Martin, A. W., & Schum, D. A. (1987). Quantifying burdens of proof: A likelihood ratio approach. *Jurimetrics Journal*, 27, 383-402.
- McCauliff, C. M. A. (1982). Burdens of proof: Degrees of belief, quanta of evidence, or constitutional guarantees? *Vanderbilt Law Review*, 35, 1260-1335.
- Michigan Supreme Court Committee on Model Criminal Jury Instructions (n.d.). *M Crim JI 1.9: Presumption of innocence, burden of proof, and reasonable doubt*. Michigan Courts. <https://www.courts.michigan.gov>
- Miller, M. K., & Bornstein, B. H. (2004). *Juror stress: Causes and interventions*. *Thurgood Marshall Law Review*, 30, 237-269.
- Mitchell, T. L., Haw, R. M., Pfeifer, J. E., & Meissner, C. A. (2005). Racial bias in mock juror decision-making: A meta-analytic review of defendant treatment. *Law and Human Behavior*, 29, 621-637. <https://doi.org/10.1007/s10979-005-8122-9>
- Mueller-Johnson, K., Dhami, M. K., & Lundrigan, S. (2018). Effects of judicial instructions and juror characteristics on interpretations of beyond reasonable doubt. *Psychology, Crime and Law*, 24, 117-133. <https://doi.org/10.1080/1068316x.2017.1394461>
- Montgomery, J. W. (1998). The criminal standard of proof. *New Law Journal*, 582-584.

- Nagel, S. S. (1979). Bringing the values of jurors in line with the law. *Judicature*, 63, 189-195.
- Nagel, S. S., & Neef, M. G. (1979). *Decision theory and the legal process*. Lexington, MA: Lexington Books.
- Nebraska Supreme Court (n.d.). *Pattern jury instructions: Criminal – Reasonable doubt*. In *Victor v. Nebraska* instruction model. <https://www.law.cornell.edu/supct/html/92-8894.ZO.html>
- Newman, J. O. (1993). Beyond “reasonable doubt.” *New York University Law Review*, 68, 979-1002. <https://doi.org/10.1093/lpr/mgm010>.
- Oregon Judicial Department (n.d.). *Oregon criminal jury instruction 3.01: Reasonable doubt*. <https://www.courts.oregon.gov/services/jury/Pages/criminal-jury-instructions.aspx>
- Park, K., Seong, Y., Kim, M., & Kim, J. (2016). Juror adjustments to the reasonable doubt standard of proof. *Psychology, Crime & Law*, 22(6), 599–618. <https://doi.org/10.1080/1068316X.2016.1168427>
- Pi, D., Parisi, F., & Luppi, B. (2019). Quantifying reasonable doubt. *Rutgers University Law Review*, 72, 455. <https://doi.org/10.2139/ssrn.3226479>
- Picinali, F. (2015). The threshold lies in the method: Instructing jurors about reasoning beyond reasonable doubt. *The International Journal of Evidence & Proof*, 19(3), 139-153. <https://doi.org/10.1177/1365712715571511>
- Power, R. C. (1999). Reasonable and other doubts: The problem of jury instructions. *Tennessee Law Review*, 67, 45-123. <https://doi.org/10.2139/ssrn.1750267>
- Reifman, A., Gusick, S. M., & Ellsworth, P. C. (1992). Real jurors' understanding of the law in real cases. *Law and Human Behavior*, 16(5), 539-554. <https://doi.org/10.1007/BF01044622>

- Scurich, N., & John, R. S. (2017). Jurors' presumption of innocence. *Journal of Legal Studies*, 46, 187-206. <https://doi.org/10.1086/690450>
- Severance, J., Greene, E., & Loftus, E. F. (1984). *Toward criminal jury instructions that jurors can understand*. *Journal of Criminal Law & Criminology*, 75(1), 198-233. <https://scholarlycommons.law.northwestern.edu/jclc/vol75/iss1/6>
- Shapiro, B. (1991). *Beyond reasonable doubt and probable cause: Historical perspectives on the Anglo-American law of evidence*. Berkeley: University of California Press. <https://doi.org/10.1525/9780520313408>
- Shapiro, J. A., & Muth, K. T. (2020). Beyond a reasonable doubt: Juries don't get it. *Loyola University of Chicago Law Journal*, 52, 1029.
- Simon, R. J. (1970). Reasonable doubt. *The Journal of Applied Behavioral Science*, 6, 203-209. <https://doi.org/10.1177/002188637000600205>
- Simon, R. J., & Mahan, L. (1971). Quantifying burdens of proof: a view from the bench, the jury and the classroom. *Law and Society Review*, 5, 319-330. <https://doi.org/10.2307/3052837>
- Solan, L. M. (2001). Refocusing the burden of proof in criminal cases. *Texas Law Review*, 78, 105-147.
- Stoffelmayr, E., & Diamond, S. S. (2000). The conflict between precision and flexibility in explaining "beyond a reasonable doubt". *Psychology, Public Policy, and Law*, 6(3), 769. <https://doi.org/10.1037//1076-8971.6.3.769>
- Tiersma, P., (2001). *The rocky road to legal reform: Improving the language of jury instructions*. *Brooklyn Law Review*, 66(4), 1081. <https://brooklynworks.brooklaw.edu/blr/vol66/iss4/4>
- Tillers, P., & Gottfried, J. (2006). Case comment—*United States v. Copeland*, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): A collateral attack on the legal maxim that proof beyond a

reasonable doubt is unquantifiable? *Law, Probability and Risk*, 5(2), 135-

157. <https://doi.org/10.1093/lpr/mgl015>

White, L. T., & Cicchini, M. D. (2019). Is reasonable doubt self-defining. *Villanova Law Review*, 64, 1.

Young, W. (2003). Summing-up to juries in criminal cases - what jury research says about current rules and practice. *Criminal Law Review*, 665-689.

Young, W., Cameron, N., & Tinsley, Y. (2001). *Juries in criminal trials* (New Zealand Law Commission Report No. 69). Wellington: New Zealand Law Commission.

Zander, M. (2000). The criminal standard of proof – how sure is sure? *New Law Journal*, 150, 1517-1519.

Cases Cited

United States v. Fatico, 1978

Victor v. Nebraska, 1994

R v Lifchus (1997) 3 SCR 320

Green v The Queen (1997) 191 CLR 334

R v Wanhalla (2007) NZCA 229

R v Majid [2009] EWCA Crim 2563

Appendix A – Examples of responses falling into each synonym for BRD (N = 217)

Category	Examples of responses
Doubt	<i>“it is fact”, “no doubt(s)”, “no question”, “there is no room for any questioning/doubt”, “no reasonable doubt”, “guilt is proved further than doubting”, “there may still be a niggle of doubt, but a very small one, which should be insignificant”, “there is still doubt involved but it relates to how likely they believe they are guilty”, “when guilt is proved further than doubting”</i>
Certainty/ probability	<i>“100%”, “absolutely”, “certain”, “100% sure”, “99%”, “99% certain”, “strongly certain that there is a high probability”, “definitely”, “almost certain”, “90% that the argument is likely”, “80% sure”, “sure”, “almost sure”, “quite sure”, “a high percentage of chance”, “greater than 50 per cent likelihood”, “a higher evidential threshold than a balance of probability”, “in the balance of probabilities”, “more likely to have committed the crime than not”, “lack of certainty”, “could not be 100% certain”, “can’t be 100% sure”, “as certain as possible”, “not be swayed on the probability”, “uncertainties”</i> LACK OF CERTAINTY OF GUILT - POINTING TO INNOCENCE: <i>“not 100% certain”, “cannot be said with certainty”, “unsure of the defendant’s guilt”, “enough uncertainty”, “slight incline they may not have done it”, “any, even slight, possibility than an individual is innocent”, “likely to be innocent on the balance of probabilities”</i>
Amount of proof/ evidence	<i>“all the evidence”, “evidence is substantial”, “the vast majority of evidence”, “there is so much evidence”, “a large amount of proven evidence”, “the majority of the evidence”, “sufficient evidence”, “enough</i>

*proof or evidence”, “the amount of proof present must be exceeded”,
“evidence/single evidence”*

LACK OF PROOF/EVIDENCE OF GUILT - POINTING TO

INNOCENCE: *“no evidence”, “no other evidence proven to make the
defendant guilty”, “not enough evidence”, “insufficient evidence”, “just
one or two pieces of evidence”, “lack of evidence”*

Strength of proof/ evidence	<i>“cast iron evidence and witnesses”, “concrete evidence”, “incontestable proof”, “strong evidence”, “proof that can be provided which clearly and unequivocally supports a given statement”, “evidence presented is fact and not circumstantial”, “evidence presented which support the verdict of guilt and refute any counter argument or evidence”, “the information gathered is strong enough”, “evidence you are shown has been properly collected and verified and delivered in a way that people understand”, “valid evidence”, “good evidence”, “relevant evidences, exhibits, alibis, witnesses”, “high level of evidence”, “slight evidence missing or evidence could be looked at as for or against but the rest of the evidence is overwhelming clear”, “clear and convincing proof”, “no concrete proof however...”</i>
Alternative explanations	<i>“points to one conclusion”, “no other alternative explanation”, “no other suspects or lines of enquiry”, “no other logical explanation”, “no realistic way it could be otherwise”, “no alternative pragmatic explanation”, “no other probable or feasible explanation”, “no meaningful argument can be put forward”, “no other reasonable explanation”</i>

Note. Quotes have been abbreviated due to space restrictions. Typos/misspellings in participants’ responses have been corrected here for readability, but remain in the raw data in the accompanying supplementary materials on OSF:

https://osf.io/ycvpd/?view_only=3991a389cdf949d79802411777c22a22

Appendix B – Further/exploratory analyses revealing participants’ conceptualization of other aspects of jury decision-making in criminal trials (N = 217)

Category	Examples of responses
Presumption of innocence	Innocent: <i>“innocent until proven guilty” “not guilty until...”</i> • Guilty: <i>“guilty until proven innocent”, “without any doubt guilty”</i>
Burden of proof	Prosecutor: <i>“prosecutor is required to prove that the person has committed the crime”, “evidence from the prosecution”, “prosecution, who must demonstrate that a defendant is guilty”</i> • Defence: <i>“defendant cannot prove that they did not commit the crime”</i>
Features of judgment process	<i>“becoming confident, in light of evidence, that the defendant is guilty”, “convinced the person has committed the crime by the evidence shown”, “not a doubt based upon sympathy or prejudice, and instead, is based on reason and common sense”, “ability to weigh arguments fairly and logically, and determine which of those arguments are most relevant for making a decision”</i>
Characteristics of the decision-maker	<i>“ordinary human being”, “reasonable person”, “ability to weigh arguments fairly and logically, and determine which of those arguments are most relevant to making a decision”, “any person with an average life experience would believe something to be true or false but could not be 100% certain”</i>

Note. Quotes have been abbreviated due to space restrictions. Typos/misspellings in participants’ responses have been corrected here for readability, but remain in the raw data available in the accompanying supplementary materials on OSF

https://osf.io/ycvdp/?view_only=3991a389cdf949d79802411777c22a22