Law and Human Behavior

Do Legal Professionals Make More Rational Decisions Than Nonprofessionals? The Biasing Effect of the Number of Verdict-Alternatives on Conviction Rates

Meir Simha Barneron, Yoram Rabin, Yaniv Vaki, and Amit Pundik

Online First Publication, October 29, 2026. https://dx.doi.org/10.1037/lhb0000682

CITATION

Barneron, M. S., Rabin, Y., Vaki, Y., & Pundik, A. (2026). Do legal professionals make more rational decisions than nonprofessionals? The biasing effect of the number of verdict-alternatives on conviction rates. *Law and Human Behavior*. Advance online publication. https://dx.doi.org/10.1037/lhb0000682

American Psychological Association
American Psychology-Law Society
Law and Human Behavior

© 2026 American Psychological Association
ISSN: 0147-7307
https://doi.org/10.1037/lhb0000682

Do Legal Professionals Make More Rational Decisions Than Nonprofessionals? The Biasing Effect of the Number of Verdict-Alternatives on Conviction Rates

Meir Simha Barneron[1], Yoram Rabin[2], Yaniv Vaki[2], and Amit Pundik[3]
[1] The Seymour Fox School of Education, The Hebrew University of Jerusalem
[2] Faculty of Law, College of Management Academic Studies
[3] The Buchmann Faculty of Law, Tel Aviv University



Objective: "Guilty" or "not guilty" is typically the choice that decision-makers face in criminal trials. However, Scotland and Israel stand out for their use of a third, intermediate, verdict (Scotland having only recently abolished its "not proven" option). In such verdicts, decision-makers may opt not to *fully* acquit the defendant—who nevertheless walks free, without criminal sanction. Rationally, the availability of intermediate verdicts should not affect the probability of conviction, since the legal question remains whether the evidence suffices to prove guilt "beyond-reasonable-doubt." If the number of verdict-alternatives *does* influence conviction rates, the decision making is distorted—an effect we term "N-verdict bias." We investigated the extent to which professional legal decision-makers are vulnerable to this bias. ***Hypotheses:*** We expected the N-verdict bias to arise among both professionals and nonprofessionals, but to be of lower magnitude among the former. ***Method:*** Seven studies were conducted ($N = 1{,}934$) on participants with varying levels of legal expertise: bar disciplinary-panel adjudicators, practicing lawyers, law students, and laypeople. In Studies 1–5, participants rendered a verdict on a fictitious murder case, under two- versus three-verdict conditions. Study 6 tested whether receiving instructions about the beyond-reasonable-doubt standard eliminated the bias. Study 7 ascertained the internal confidence-threshold participants deemed necessary for conviction under two- versus three-verdict conditions. ***Results:*** The N-verdict bias arose among laypeople, law students, *and* professionals: The availability of an intermediate verdict decreased the conviction rate. This bias was uninfluenced by legal expertise and was present even when participants were instructed to convict *only* if guilt were proven beyond-reasonable-doubt. The confidence-threshold required for conviction did not vary significantly between the two conditions. ***Conclusions:*** Defendants are less likely to be convicted in a three-verdict system, making the number of verdict-alternatives another source of bias in legal decision making that legal expertise does not immunize against.

Public Significance Statement
Worldwide, jurisdictions maintain the ideal that conviction decisions should be based on the evidence alone and not be influenced by other factors, such as the number of verdict-alternatives. However, our findings reveal a bias whereby adding a third (intermediate) verdict vis-à-vis guilty/not guilty dramatically reduces the conviction rate, even when the evidence and charges remain identical. Crucially, the data indicate that professional legal decision-makers are as prone to this bias as laypeople are. Policymakers should take these findings into account when considering reforming the verdict-alternatives presented to legal decision-makers.

Keywords: rationality, expertise, law, judgment, social decision making

Supplemental materials: https://doi.org/10.1037/lhb0000682.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Liana C. Peter-Hagene served as Action Editor.
Meir Simha Barneron https://orcid.org/0000-0002-2257-0173
Yoram Rabin https://orcid.org/0000-0003-3058-058X
Yaniv Vaki https://orcid.org/0000-0003-1503-2235
Amit Pundik https://orcid.org/0000-0001-5835-9073

Five of the studies were preregistered, as follows: Study 2, https://aspredicted.org/1LR_TSH; Study 4, https://aspredicted.org/9DZ_SBT; Study 5, https://aspredicted.org/nz56-6j38.pdf; Study 6, https://aspredicted.org/57x2-m4jd.pdf; and Study 7, https://aspredicted.org/pmdb-48tg.pdf. The data collected, as well as the R code used for data processing and all analyses, are available at https://data.mendeley.com/datasets/ck3452sb5v/2. This article is based on data presented at the 2025 Conference on Empirical Legal Studies (October 24–25, Georgetown University Law Center, Washington, DC). In preparing this article, the authors used an AI assistant (Claude, Anthropic) for language editing and for verifying compliance with APA style; the authors

continued

"Guilty" or "not guilty" is the choice that decision-makers face in criminal trials in virtually all countries. However, Israel and Scotland stand out as singular exceptions, due to their use of three-verdict systems. In Israel, judges can select from "guilty," "full acquittal," or "acquittal for the benefit of the doubt" (henceforth, "ABD"). And Scotland has only recently switched to a two-verdict system, having employed a third verdict—that of "not proven"—for hundreds of years, until its abolition in January 2026. While such verdicts are perceived as offering decision-makers an intermediate alternative in criminal trials, the crucial point is that the defendant is acquitted and walks free without any criminal sanction (Barbato, 2005; Bray, 2005; Phalen, 2018; Rabin & Vaki, 2021).

Neither the Scots "not proven" nor the Israeli ABD verdict was created by the legislature. Rather, each originated in, and was later developed by, the courts (Rabin & Vaki, 2021). Against that backdrop, the long-standing legal debate on the pertinence of these alternatives (henceforth: "intermediate verdicts") has recently been attracting renewed theoretical interest, together with public action to abolish them. On the one hand, in the first book-length theoretical study of intermediate verdicts, Picinali (2022) relied on decision theory to justify this alternative as viable under a wide range of theories of punishment. The Scottish Parliament, on the other hand, recently abolished the "not proven" verdict in its Victims, Witnesses, and Justice Reform (Scotland) Act 2025, following a major consultation exercise (Scottish Government, 2021; see also Chalmers et al., 2021).

Both of these intermediate verdicts have been widely used: Before its abolishment, Scotland's "not proven" was used in 20% of jury acquittals, rising to 44% of the acquittals in rape and attempted-rape cases (Scottish Government, 2021). In Israel, our own previous research shows that ABD verdicts have been used even more frequently, corresponding to 48.1% and 49.5% of acquittals in the District and Magistrates' Courts, respectively, rising to 56.8% in sexual offenses (Rabin et al., 2025). A notable difference between the "not proven" and the ABD verdicts concerns the formal legal implications of such an acquittal for the defendants. While the Scots "not proven" had no formal ramifications for defendants, the Israeli ABD has several: Among others, it might deny them compensation for wrongful prosecution and can be used against them in child custody disputes and permanent residency applications (Rabin & Vaki, 2021).

The present research foregrounds the issue of intermediate verdicts to explore a central question at the heart of criminal trials: whether and how judicial expertise helps in achieving rational decision making. More specifically, we set out to empirically investigate *the impact of intermediate verdicts on the decisions rendered by legal professionals* in criminal trials. It is crucial to study legal professionals because, in most jurisdictions around the globe, criminal verdicts are rendered by professional judges rather than by juries consisting of laypeople. Furthermore, even in jury-based jurisdictions, professionals serve as adjudicators in various contexts where considerable sanctions can be imposed on individuals (such as disciplinary proceedings or administrative offenses). Assessing whether, and to what extent, legal and judicial expertise contributes to rational decision making is, therefore, of immense practical value. However, to date, the empirical studies on intermediate verdicts surveyed above have only analyzed the behavior of lay decision-makers—that is, jury members with no legal background. The only extant study conducted on a non-lay sample investigated the views of Scottish legal professionals on the "not proven" verdict (Curley et al., 2021). Given this lacuna, the present research addressed the following research questions: (a) Does the availability of an intermediate verdict reduce conviction rates among professional decision-makers? And (b), if so, does legal expertise attenuate this effect?

A Rational Perspective on Intermediate Verdicts

Criminal trials revolve around a single question: Is the defendant guilty of the crime(s) of which they are accused? To answer that question, legal decision-makers must consider the evidence presented before them and decide whether it is sufficient to prove the defendant's guilt. According to the rational model of legal decision making, decision-makers should convict the defendant only if the evidence is sufficiently strong to satisfy the standard of proof required by the law (such as beyond-reasonable-doubt). In England and Wales, the formal *Jury Manual* states, "The jury's tasks are to weigh up the evidence [and] decide what has been proved and what has not … The jury must try the case only on the evidence and arguments they hear in court" (Judicial College, 2026, pp. 3-3). A similar instruction exists in the United States (Judicial Council of the United States Eleventh Judicial Circuit, 2025). As Curley et al. (2021, p. 263) noted, "The role of a jury is not to decide on whether

reviewed all suggested edits and take full responsibility for the content of the article. The authors have no known conflicts of interest to disclose. This research was supported by the Israel Science Foundation (Grant 366/24 awarded to Amit Pundik). The authors are grateful to Dalit Mansoor and Romi Alimi for their research assistance.

Meir Simha Barneron played a lead role in formal analysis, software, visualization, and writing–original draft, a supporting role in resources, and an equal role in conceptualization, data curation, funding acquisition, investigation, methodology, project administration, supervision, validation, and writing–review and editing. Yoram Rabin played a lead role in investigation and resources, a supporting role in visualization, and an equal role in conceptualization, data curation, funding acquisition, methodology, project administration, supervision, validation, and writing–review and editing. Yaniv Vaki played a lead role in investigation and resources and an equal role in conceptualization, data curation, funding acquisition, methodology, project administration, supervision, validation, visualization, and writing–review and editing. Amit Pundik played a supporting role in formal analysis, resources, and writing–original draft and an equal role in conceptualization, data curation, funding acquisition, investigation, methodology, project administration, software, supervision, validation, visualization, and writing–review and editing.

The data are available at https://data.mendeley.com/datasets/ck3452sb5v/1.

The experimental materials are available at https://data.mendeley.com/datasets/ck3452sb5v/1.

The preregistered design is available at Study 2, at https://aspredicted.org/1LR_TSH; Study 4, at https://aspredicted.org/9DZ_SBT; Study 5, at https://aspredicted.org/nz56-6j38.pdf; Study 6, at https://aspredicted.org/57x2-m4jd.pdf; and Study 7, at https://aspredicted.org/pmdb-48tg.pdf.

Correspondence concerning this article should be addressed to Meir Simha Barneron, The Seymour Fox School of Education, The Hebrew University of Jerusalem, Mount Scopus Campus, Jerusalem 9190501, Israel. Email: meir.barneron@mail.huji.ac.il

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

an accused did factually commit the crime. Rather, their job is to use the evidence to establish whether the prosecution has proved their case beyond reasonable doubt."

Notably, while the strength of the evidence may be scalar, its sufficiency is binary: Is the evidence sufficient or not? If the evidence suffices, rational decision-makers should convict, regardless of the number of verdict-alternatives. In principle, then, if there is only one alternative verdict to conviction, they should convict because the evidence suffices to prove the defendant's guilt. If there are two additional alternative verdicts to conviction, they should also convict (because, again, the evidence suffices to prove the defendant's guilt). In light of this legal principle, the number of alternatives is irrelevant to the rational decision of whether to convict.

The only effect that the availability of an intermediate verdict may have without contradicting the requirements of the rational model is on the distribution between the nonconviction alternatives. In a two-verdict system, all the decisions *not* to convict should lead rational decision-makers to render the alternative verdict (i.e., to acquit). By contrast, in a three-verdict system, deciding not to convict can lead decision-makers to opt for either full acquittal or an intermediate verdict, and the rational model does not require a certain distribution between the two. In rational decision making, any intermediate verdict should, hence, come at the expense of the other non-convicting alternative, not at the expense of a "guilty" verdict.

Consider the following hypothetical. Numerous judges are assigned to the same trial under identical conditions and then asked, separately, to render a verdict under a two-verdict system: convict or acquit. Suppose 60% convict and 40% acquit (see the two-verdict setting in Table 1). These outcomes indicate that the 60% were satisfied beyond-reasonable-doubt that the evidence sufficed to convict. The other 40% found it insufficient, either because they believed the defendant to be innocent or because, though they believed them to be guilty, this belief was clouded by a reasonable doubt that the evidence did not dispel.

Imagine now an alternative scenario. Different judges are assigned to the aforementioned trial, again under identical conditions, but must render a verdict in a three-verdict system: convict, fully acquit, or acquit on the benefit of the doubt (Israel's ABD). They face the same legal question as before: whether the evidence suffices to prove the defendant's guilt beyond-reasonable-doubt. If they comply with the rational model of legal decision making, the conviction rate should, ostensibly, remain at 60%, since both the evidence and the standard of proof are identical to those of the previous scenario.

The rational model allows various distributions of the two acquittal verdicts. The 40% acquittal rate might split, for instance, into 30% full acquittals and 10% ABD (see Setting A in Table 1, which compares hypothetical outcomes), or 10% full acquittals and 30% ABD (Setting B), or even 0% full acquittals and 40% ABD (Setting C). What A–C share is that the ABD was selected exclusively at the expense of full acquittals: While the full acquittal rate varies, the conviction rate holds at 60%. The judges' decision making thus accords with the rational model, because they answered the legal question based solely on their assessment of the evidence and its sufficiency to prove guilt beyond-reasonable-doubt—that is, their conviction rate was unaffected by the number of acquittal alternatives. Note that judges can comply with the rational model without rendering any intermediate verdicts at all: In Setting D (Table 1), none opts for the ABD, and the split remains 60% conviction and 40% acquittal.

Now consider Settings E and F (Table 1), where judges do *not* follow the rational model: Here, the ABD was selected, at least partly, at the expense of convictions. In Setting E, the ABD reduced the conviction rate to 50%, while full acquittals held at 40%; in Setting F, it reduced both. In these two settings, the distribution reveals that judges answered the legal question differently when the ABD was available, compared to when it was not.

The N-Verdict Bias

To the extent that the conviction rate is affected by the variable "number of verdict-alternatives," this effect constitutes a systematic deviation from a rational norm—that is, a *bias*. In this article, we term this the "N-verdict bias" because, as we have seen, per the rational model, the probability of conviction should be determined solely by the evidence and its ability to meet the required legal threshold. If one's decision to convict is at all influenced by the *number* (N) of alternative verdicts on offer, one is biased by a factor that is neither the evidence nor the standard of proof.

A priori, one might consider the N-verdict bias to be simply one more scenario where generating more options is *not* beneficial (see Guthrie, 2002, for various such cases in the context of negotiation). However, it is important to emphasize that this bias does not necessarily result from *adding* an intermediate verdict. It may instead result from *limiting* the number of verdicts to two. The criminal justice system aspires to balance the risks of false positives (wrongful convictions) and false negatives (erroneous acquittals) according to the relative social costs of each. It is unclear whether the optimal balance between these risks is achieved by the conviction rate obtained under a two- or a three-verdict system (see the discussion in the Should Intermediate Verdicts Be Encouraged or

Table 1
Hypothetical Example of Distribution of Verdicts Under a Two- Versus Three-Verdict System

System	Setting	Conviction (%)	Full acquittal (%)	ABD (%)	Rational?
Two-verdict		60	40		
Three-verdict	A	60	30	10	Yes
Three-verdict	B	60	10	30	Yes
Three-verdict	C	60	0	40	Yes
Three-verdict	D	60	40	0	Yes
Three-verdict	E	50	40	10	No
Three-verdict	F	50	5	45	No

Note. ABD = acquittal for the benefit of the doubt.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Avoided? section). Either way, the fact that the conviction rate differs between the two systems constitutes a bias because the decision to convict is influenced by the number of verdict-alternatives. It is for this reason that we do not label it "intermediate-verdict bias" or "no-intermediate-verdict bias."

The N-Verdict Bias in the Literature

The extant research dealing with intermediate verdicts focuses exclusively on the Scots "not proven"; no experiments have been conducted on the Israeli ABD to date. Previous studies found that the availability of "not proven" reduced the rates of both "guilty" and "not guilty" outcomes. The most comprehensive empirical study on the "not proven" verdict was commissioned by the Scottish Government (Ormston et al., 2019). It remains the largest mock jury study ever conducted in the United Kingdom, involving 64 mock juries and 969 participants, who watched and deliberated on 2-hr-long filmed trials, these being deliberately finely balanced. The findings revealed that the availability of "not proven" rendered individual jurors significantly less likely to convict, both before and after deliberation: 28% of jurors in the three-verdict condition favored "guilty" in predeliberation, compared to 38% in the two-verdict condition; postdeliberation, the figures were 22% versus 31%, respectively, suggesting an effect on jurors' verdict preferences, independent of group discussion. The study also found that "[t]here were no statistically significant differences in the length of time taken to reach a verdict between juries asked to choose between two and three verdicts" and that "[t]he presence or absence of the not proven verdict had little impact on the length of deliberations" (Ormston et al., 2019, p. 30 and p. viii, respectively). Jackson et al.'s (2024) quantitative meta-analysis, pooling data from 10 mock trials involving 1,778 jurors, found that the rate of conviction was approximately 40% lower under the Scots three-verdict system compared to the conventional two-verdict system, with conviction rates (pooled across trials) of 39.2% under the two-verdict versus 28.7% under the three-verdict system.

Furthermore, the availability of "not proven" reduces not only the conviction rate but also the proportion of "not guilty" verdicts. Smithson et al. (2007) found that "not proven" primarily drew from "not guilty" verdicts. Similar findings were presented by Hope et al. (2008). For example, in their Study 1 (sexual assault transcript), verdict distributions differed sharply between systems: In the two-verdict condition, 39% returned "not guilty," while in the three-verdict condition, only 7% did so. The aforementioned meta-analysis by Jackson et al. (2024) also found that the frequency of "not guilty" was 88% lower under Scotland's three-verdict system.

The N-Verdict Bias and the Possible Psychological Mechanism Behind It

A number of studies have analyzed this bias (albeit labeling it in different ways), as part of the ongoing endeavor to isolate the psychological mechanism at its root. One prevalent explanation is the so-called "compromise effect," whereby individuals facing uncertainty avoid extreme options in favor of a middle ground (Simonson, 1989; see also Barneron & Yaniv, 2020). In a two-verdict system, "guilty" and "not guilty" are clearly opposites; but as Jackson et al. (2024) and Curley, Murray, et al. (2022) observed, the availability of a "not proven" verdict may heighten further the sense that the remaining two verdicts are polarized (as extremes), rendering the intermediate option more attractive. Evidence that decision-makers view "not proven" as a middle-ground option can be found in studies concluding that jurors who choose "not proven" hold subjective beliefs of culpability (measured by the probability they assigned to the defendant's potential guilt) that are significantly stronger compared to those deciding "not guilty" and weaker compared to those deciding "guilty" (Smithson et al., 2007). Proponents of the compromise-effect mechanism draw support from the fact that the "not proven" group in Smithson et al.'s (2007) study reported higher decisional difficulty and internal conflict, suggesting that the third verdict may enable people to resolve cognitive dilemmas where doubt exists about the defendant's guilt, without committing to conviction. Understood thus, the mechanism would appear to operate through extremeness aversion: When decision-makers' internal beliefs sit in an ambiguous zone, the three-verdict system provides a course of action that aligns more comfortably with that position, with the result that fewer of them gravitate toward to conviction.

Another mechanism proposed in the literature involves semantic and cognitive reframing. The language used to delineate verdict options has been found to significantly influence juror cognition, shifting their focus from moral judgment to evidentiary standards. Jackson et al. (2024), for instance, described a distinction between "truth" and "proof": Terms such as "proven" can encourage jurors to focus on weaknesses in the prosecution's evidence because the burden lies with the prosecution. According to these authors, "not proven" may shift torn jurors toward acquittal by reminding them that, without proof beyond-reasonable-doubt, they should acquit even if they happen to believe the defendant is guilty.

On this premise, in *borderline* cases, if "not guilty" is experienced psychologically as "declaring innocence," some jurors may be reluctant to acquit in a two-verdict system and may therefore convict instead. In contrast, providing them with a third alternative of "not proven" could enable them to acquit without making a declaration of innocence. Individually, this may help them resolve their psychological friction, but more importantly, it potentially reduces convictions in borderline evidence cases, thereby explaining the difference in conviction rates between the two- and three-verdict systems.

Along similar lines, Curley et al. (2019) suggested that, in practice, jurors may treat "not guilty" as something resembling "proof of innocence," whereas "not proven" can be understood as indicating the prosecution's failure to meet the burden of proof. In Curley, Munro, et al. (2022), the authors compared a two-verdict system that was based on "guilty/not guilty" with a two-verdict system based on "proven/not proven," finding that the latter yielded significantly fewer convictions. The authors suggested that the "not proven" wording anchors jurors to the burden of proof and the state's failure to meet it, rather than the defendant's factual innocence. By highlighting the distinction between "Did they do it?" (truth) and "Did the state prove they did it?" (proof), the authors found that the "guilty" verdict requires a higher subjective probability of guilt than a verdict of "proven." They also suggested that the term "guilty" conveys moral condemnation and may trigger intuitive, emotional decision making, whereas "proven" prompts jurors to think analytically about whether the standard of proof has objectively been met.

All of the foregoing findings raise an interesting question about the potential effect of the number of verdicts on the decision-maker's internal confidence-threshold (the level of confidence in the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

defendant's guilt that must be reached in order for them to opt for conviction). One might hypothesize that the N-verdict bias occurs because the additional verdict increases the internal confidence-threshold regarding the defendant's guilt. If this is, indeed, the mechanism at play, when the number of verdicts is increased from two to three and the individual is offered an intermediate option to "acquit on the benefit of the doubt," fewer people will choose to convict under the three-verdict condition because the internal degree of confidence that the individual deems necessary to convict is increased as a result. This, too, was tested in our studies.

The N-Verdict Bias and Legal Expertise

The central question of the present research is whether the N-verdict bias is attenuated by the decision-maker's having legal expertise or whether legal professionals are as vulnerable to the bias as nonprofessionals. There are reasons to expect a difference in the conviction rate under two- versus three-verdict conditions among legal professionals, but also reasons to expect this difference to be tempered, in contrast to the case of laypeople. On the one hand, legal fact-finding does not necessarily provide the conditions under which expertise reliably eliminates all errors of judgment. General theories of expertise suggest that skill-based intuition develops primarily in environments that are sufficiently repetitive and provide rapid, unambiguous feedback on one's decisions. But criminal fact-finding does not fit this mold: It often involves contested human behavior, uncertain evidence, and little direct feedback about whether the decision was correct (Kahneman & Klein, 2009; Shanteau, 1992; Tetlock, 2005). Consistent with this view, studies on judicial contexts have shown that professional legal decision-makers remain susceptible to cognitive biases, including anchoring, hindsight bias, and other intuitive distortions (Guthrie et al., 2001; Rachlinski & Wistrich, 2018). Thus, there is no reason to assume that legal expertise will immunize decision-makers against the N-verdict bias.

On the other hand, legal expertise may still reduce the magnitude of such biases. Legal training teaches professionals to distinguish admissible from inadmissible evidence, organize the admissible evidence according to legally relevant factual questions, reason in terms of burdens and standards of proof, and focus on objective evidentiary sufficiency rather than on general impressions of culpability (Schauer & Spellman, 2017). Empirical findings support the view that legal expertise goes at least some way toward insulating decision-makers from subjectivity. Teichman et al. (2023), for example, found that outcome bias and anti-inference bias affected laypeople, law students, prosecutors, and defense lawyers similarly. By contrast, an irrelevant numerical anchor influenced laypeople and law students in their decision making, whereas experienced legal professionals were unaffected. Accordingly, our hypothesis was moderate: We expected legal professionals to exhibit some degree of N-verdict bias, but to a lesser extent than law students and laypeople.

The Present Studies

We conducted seven studies, involving a total of 1,934 participants. In Studies 1–5, we tested whether availing of an intermediate verdict affected the conviction rate, comparing three different populations: (a) legal professionals, our primary (reference) group of interest (Studies 1 and 2); (b) law students (Studies 3 and 4); and (c) laypeople (Study 5). Participants were asked to read a summary of the hypothetical case of a man accused of murder and then to render a verdict either in a two-verdict or in a three-verdict system. These five studies focused on determining whether the availability of an intermediate verdict generated the N-verdict bias among legal professionals and whether legal expertise attenuated this bias.

We also conducted two further studies, looking into the mechanism underlying the N-verdict bias itself. In the first of these additional studies (Study 6), we examined the impact of giving instructions to lay participants on the beyond-reasonable-doubt standard. In short, we tested whether the N-verdict bias persisted even when participants were instructed not to convict unless there was *no reasonable doubt* regarding the defendant's guilt. It was important to test this specifically because one might hope that reminding jurors of the required standard would focus their attention on the very question that the rational model of legal decision making enshrines. Indeed, we expected to find that adding such an instruction would neutralize the N-verdict bias. Finally, in Study 7, we tested whether the mere presence of an intermediate verdict would elevate the level of confidence that laypeople required in order to convict a defendant. To do this, we measured the influence of the number of verdicts on participants' opinions about the internal confidence-threshold they would need to reach in order to convict someone of murder.

Research Transparency Statement

All seven studies received the prior formal approval of either the Ethics Committee of Tel Aviv University or that of the College of Management Academic Studies. For Study 1, we also obtained the approval of the National President of the Israeli Bar's Disciplinary Panels. All materials used in our studies are shown in the Supplemental Materials. The data collected, as well as the R code used for data processing and all analyses, are available at https://data.mendeley.com/datasets/ck3452sb5v/2. Studies 2, 4, 5, 6, and 7 were preregistered (see later sections).

Studies 1–5: The N-Verdict Bias and Legal Expertise

Studies 1–5 pursued two aims, aligned with our research questions and hypothesis. The first was to investigate the N-verdict bias among three distinct populations with different levels of legal expertise, as noted earlier: professional lawyers (Studies 1 and 2), law students (Studies 3 and 4), and laypeople (Study 5). Among each population, we tested the hypothesis that, when an intermediate verdict was made available, the conviction rate would be lower than when only two verdicts were made available, *across all three expertise levels*. The second aim was to investigate whether legal expertise attenuated any N-verdict bias identified, by testing whether professional lawyers were less affected by the bias compared to (a) law students and (b) laypeople.

All five studies involved a similar set of materials, procedures, data processing, and statistical analysis while differing with respect to the participants and their expertise. Three of these studies were preregistered: Study 2, https://aspredicted.org/1LR_TSH; Study 4, https://aspredicted.org/9DZ_SBT; and Study 5, https://aspredicted.org/nz56-6j38.pdf.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Method

Participants

Ideally, the optimal approach to testing our hypothesis would have been to examine the effect of the availability of an intermediate verdict not only on practicing lawyers but also on a sample of professional judges. Unfortunately, setting up experiments involving presiding judges is notoriously difficult. For example, in Israel, where our studies were administered, Section 34(a) of the Judiciary Ethical Rules 2007 permits presiding judges to participate in research only with the approval of the President of the Supreme Court, and this approval is all but impossible to obtain. In 2016, the Judiciary Ethics Committee expanded this requirement to retired judges (Decision A/62/16).

To mitigate this limitation, we conducted our studies among other relevant population profiles. First of all, to ensure our sample characteristics resembled those of judges as closely as possible, we recruited lawyers who sit on the Israeli Bar's disciplinary panels. Among the populations that *are* permitted to participate in academic research, we believe this profile provides the best fit because the remit of members of these panels involves responsibilities similar to those of judges: They must make factual determinations, render verdicts of guilty versus innocent, and may impose various serious sanctions on qualified lawyers. Our participants also included other populations of legal professionals: criminal lawyers from the Prosecution Service, public and private defense lawyers, and practicing lawyers from other fields. Investigating those populations has the additional advantage of extending the practical importance of our research, given that these practitioners often make decisions with direct implications for defendants (such as whether to prosecute, whether to accept a plea bargain, and so on).

To test our hypothesis, we compared the results regarding the N-verdict bias among professionals to the results derived from two further populations: law students and laypeople. Undergraduate law students should not be considered laypeople: Their level of legal knowledge and expertise can be assumed to be greater than that of lay decision-makers, and it can also be presumed to be lower than that of legal professionals. To complete the comparison, then, as well as the sample of students, we also recruited a sample of laypeople with no legal background. By carefully calibrating our samples to represent populations with legal expertise, those with *some* expertise, and those with none, we were able to investigate the impact of this variable.

Study 1 was conducted on 29 lawyers ($M_{age} = 47.93$, $SD = 10.24$; majority, men: 65.52%) who were members of the Bar disciplinary panels in Israel's Central Area, recruited during a professional training session. Study 2 was conducted on 186 lawyers practicing in Israel ($M_{age} = 45.05$, $SD = 9.93$; majority, men: 59.68%), recruited from separate professional training sessions and meetings. Study 3 ($N = 161$; $M_{age} = 23.54$, $SD = 4.29$; majority, women: 60.87%) and Study 4 ($N = 163$; $M_{age} = 25.24$, $SD = 4.96$; majority, women: 57.67%) were conducted on students recruited from Israeli undergraduate law degrees (specifically, from large-group mandatory courses on core content). For Study 5, we recruited 501 participants from the United States ($M_{age} = 46.06$, $SD = 16.18$; majority, women: 51.30%) via the Prolific online research platform, representing the U.S. lay population. The initial sample of 626 was reduced after excluding 125 individuals (19.97%) who failed three comprehension checks, as per the preregistration specifications.

The recruitment setting, language, duration, and compensations differed across studies. In Studies 1–4, participants (who were all Hebrew speakers) took part voluntarily in a 15-min study; in Study 5, participants (who were all American English-speaking) completed a 12-min study in return for approximately $2.70 (a rate set automatically by the platform). We opted to recruit participants from the United States since there is no online platform suitable for recruiting such large samples of laypeople in Israel.

Procedure

Studies 1–4 were conducted in person, and we used the same materials for all four of them. At the outset, a three-page document providing all instructions and materials on the hypothetical case was handed to each participant. This procedure ensured that the entire procedure was self-paced and did not require any intervention from the research team (apart from distributing the questionnaires and collecting them at the end). Study 5 was conducted online. Following their consent, participants were redirected to an external platform (Qualtrics), where specific instructions were given.

Participants were asked to read the materials carefully. Their task then consisted of rendering a verdict on the case, as if they were the judge, by ticking the relevant box on a paper response sheet (Studies 1–4) or selecting the relevant verdict on-screen (Study 5) to indicate their decision. Finally, participants were required to respond to two sociodemographic questions to indicate their age and gender.

Materials

In Studies 1–5, the materials comprised a summary of the hypothetical case of a man accused of murder. According to the prosecution's case, the defendant intentionally killed another man following a verbal altercation. The materials summarized the undisputed facts, the available evidence, and the prosecution and defense claims. In Studies 1–4, the materials were in Hebrew. In Study 5, we used an English translation of the same case to make it accessible to the U.S. lay sample. The only differences between the materials used in Studies 1–4 versus Study 5 were details that were irrelevant to the facts in dispute, namely the following: the names of the parties and witnesses, the name of the park where the incident took place, the phone number of the emergency services, and the name of the police force. We changed these details in Study 5 merely to adapt the case to the U.S. context.

Participants in Studies 1–4 were presented with the following case (translated here from Hebrew and summarized).

Facts. On July 17, 2022, the deceased went for a walk in Hayarkon Park, Tel Aviv, and did not return home. The next day, police found his body in a thicket. He had died of suffocation due to respiratory failure caused by repeated blows to the head and neck with a blunt object. After the incident was reported publicly, a neighbor informed police that they had seen the defendant return from a bike ride with facial scratches and discard a bicycle pump. He was arrested and charged with murder.

Prosecution. The State alleges that the defendant, cycling on a pedestrian path, quarreled with the deceased over the right to ride there; beat him with a blunt object, thereby causing his death;

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

dragged the body into the bushes; and fled. The evidence submitted by the prosecution is that (a) Witness 1 reported seeing a fight involving a cyclist but, in questioning, admitted it was dark and could identify no one; (b) cell-network data placed the defendant within 30 m of the body around the time of death; (c) a late-night call was made from the defendant's cellphone seeking "advice" from a defense lawyer; (d) the defendant had fresh scratches, was limping, gave a shifting account of his route, and failed to produce his jacket or bicycle, eventually refusing to answer; and (e) mud on his shoes was shown forensically to match soil at the scene.

Defense. The defense argues that the State has not proved identity. The evidence submitted by the defense is that (a) Witness 1 never identified the defendant or described any violence; (b) his proximity to the crime scene would be shared by anyone in the park; (c) the scratches were caused by low-hanging branches, common among cyclists; (d) the jacket was carelessly left at a water fountain, the bicycle likely stolen, and the pump simply discarded as it was faulty; and (e) the mud is unremarkable because the park's soil is uniform throughout.

The Supplemental Materials document contains this case in full, together with the instructions delivered to participants, as well as the adaptation for a U.S. audience used in Study 5.

Design

Participants were randomly assigned to one of two conditions, as follows: In the *two-verdict condition*, they were asked to express their decision on the specific case presented to them by selecting either "to convict the defendant" or "to acquit the defendant." In the *three-verdict condition*, participants were asked to express their decision by choosing whether "to convict the defendant," "to acquit the defendant on the benefit of the doubt," or "to fully acquit the defendant."

Results

This section examines the results of Studies 1–5 from the following perspectives: (a) the percentages for the selection of each type of verdict, for each study, in the two- and three-verdict conditions; (b) the results of analyses evidencing the N-verdict bias among professionals (Studies 1 and 2), law students (Studies 3 and 4), and laypeople (Study 5); (c) the finding, contrary to our initial expectation, that legal expertise does not attenuate the N-verdict bias; and (d) the impact of ABD verdicts on full acquittals. The preregistered analyses in Studies 2, 4, and 5 were logistic regressions, comparing the conviction rates between two- and three-verdict conditions. The other analyses were not preregistered.

Table 2 presents the percentage of participants selecting each type of verdict, by condition, for Studies 1–5.

The Selection of Intermediate Verdicts

Analysis of the results of Studies 1–5 showed that, when the opportunity to select an intermediate verdict was made available, a large portion of decision-makers seized it. As shown in Table 2, in the three-verdict condition, the intermediate option accounted for between 41.67% and 72.04% of the verdicts selected, across *all* populations, including legal professionals. These findings raised the question of whether the selection of the intermediate verdict in the three-verdict condition was at the expense of full acquittal (as the rational model permits, though does not require) or "guilty" verdicts (which it proscribes).

Table 2
Results of Studies 1–5: Distribution (%) of Verdicts by Condition

Condition	*N*	Conviction (%)	Acquittal (%)	ABD (%)
Study 1				
Two-verdict	16	37.50	62.50	
Three-verdict	13	7.69	23.08	69.23
Study 2				
Two-verdict	93	24.73	75.27	
Three-verdict	93	13.98	13.98	72.04
Study 3				
Two-verdict	80	65.00	35.00	
Three-verdict	81	46.91	3.70	49.38
Study 4				
Two-verdict	83	51.81	48.19	
Three-verdict	80	26.25	3.75	70.00
Study 5				
Two-verdict	261	63.98	36.02	
Three-verdict	240	55.00	3.33	41.67

Note. ABD = acquittal for the benefit of the doubt.

If ABD verdicts were reached only at the expense of full acquittal verdicts, one might expect the full acquittal rate to be lower in the three-verdict condition than in the two-verdict condition, but not expect the conviction rate to decrease between these conditions. However, as Table 2 shows, the selection of ABD verdicts was made at the expense of both full acquittal *and* conviction verdicts. The next section formally assesses the bias created by changing the number of verdict-alternatives.

The N-Verdict Bias

To assess the existence and magnitude of the N-verdict bias among the different populations, we compared the conviction rate in the two- versus the three-verdict conditions among each population. In all studies, participants' verdicts were recoded "1" for conviction and "0" for acquittal, full acquittal, and ABD. Our dependent variable was, therefore, binary, reflecting whether the participant chose to convict (1) or not (0). The data were analyzed with both logistic regressions and chi-square tests. In the former, the independent variable was the condition to which the participant was assigned (coded "0" and "1" for the two- and three-verdict conditions, respectively).

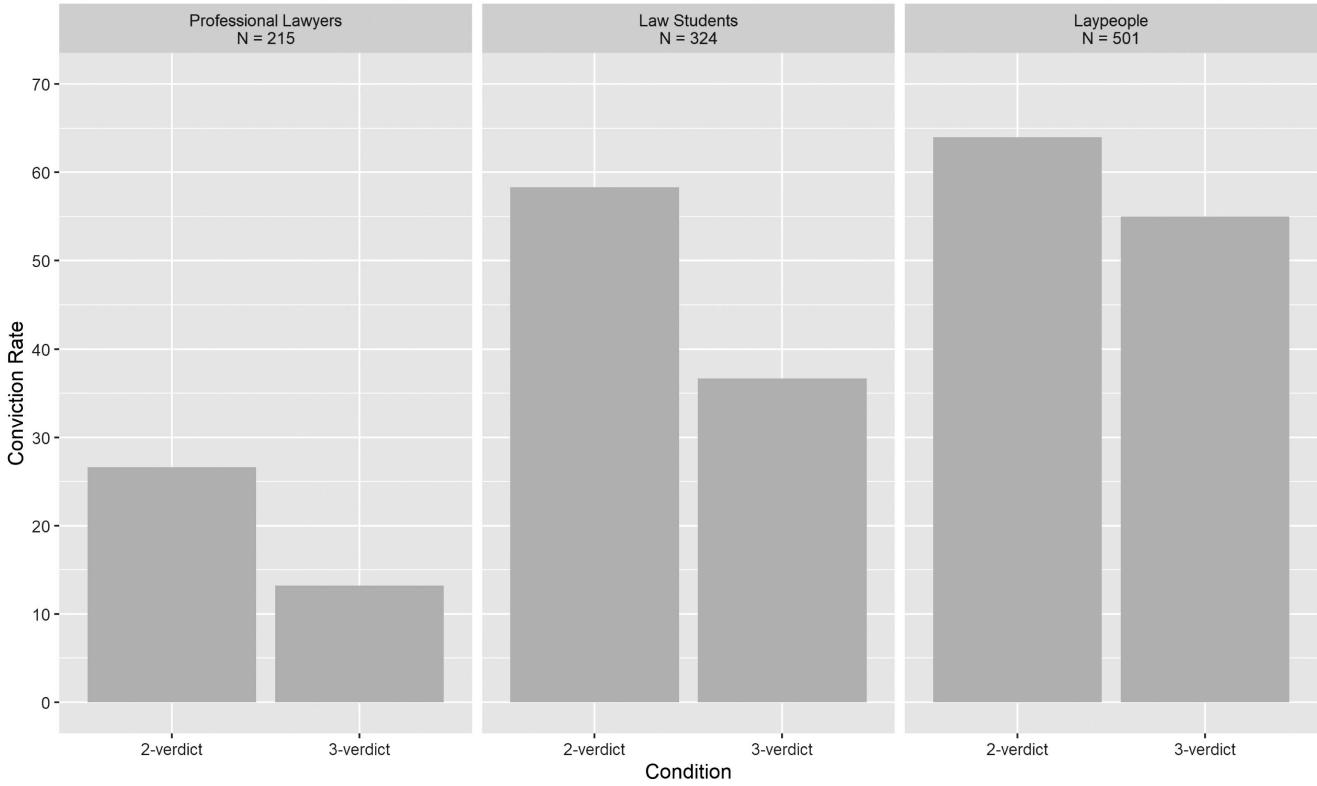
Figure 1 shows the conviction rate in the two-verdict versus the three-verdict condition, among professional lawyers (Studies 1 and 2), law students (Studies 3 and 4), and laypeople (Study 5).

As shown in Figure 1, in all three populations, the conviction rate in the three-verdict condition was lower than in the two-verdict condition. Consider, first, the population of professional lawyers. Here, we combined the data from Studies 1 and 2, giving 215 participants. Combining these two groups is justified because all the participants in both groups were legal professionals (given that all members of the Bar disciplinary panels must be qualified lawyers). The results indicated that the conviction rate dropped from 26.61% in the two-verdict condition to 13.21% in the three-verdict condition (see Figure 1). The logistic regression analysis revealed that this difference was statistically significant: $b = -0.87$, $OR = 0.42$,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 1

Conviction Rate in the Two- Versus Three-Verdict Conditions, in Each Population of Studies 1–5



$z = -2.41, p = .016$, 95% CI [0.20, 0.84]. Similar outcomes emerged when using a chi-square test: $\chi^2(1, N = 215) = 6.15$, $p = .014$.

The difference in the conviction rate between the two conditions in each study separately was not statistically significant. In Study 1 ($N = 29$), the conviction rate dropped from 37.50% in the two-verdict condition to 7.69% in the three-verdict condition (see Table 2): $b = -1.97$, $OR = 0.14$, $z = -1.70$, $p = .089$, 95% CI [0.01, 1]; $\chi^2(1, N = 29) = 3.48, p = .062$. In Study 2, the conviction rate dropped from 24.73% in the two-verdict condition to 13.98% in the three-verdict condition (see Table 2): $b = -0.70$, $OR = 0.49$, $z = -1.84$, $p = .067$, 95% CI [0.23, 1.04]; $\chi^2(1, N = 186) = 3.44$, $p = .063$.

Consider, next, the results of the analysis involving law students. First, we combined data from Studies 3 and 4, giving 324 participants. Here, too, we found that the conviction rate dropped, from 58.28% in the two-verdict condition to 36.65% in the three-verdict condition (see Figure 1). The logistic regression analysis revealed this difference to be statistically significant: $b = -0.88$, $OR = 0.41$, $z = -3.87$, $p < .001$, 95% CI [0.26, 0.65]. Similar outcomes were derived when the chi-square test was used: $\chi^2(1, N = 324) = 15.20$, $p < .001$.

The analyses for each study in its own right also revealed significant differences between the two conditions. In Study 3, the conviction rate dropped from 65.00% to 46.91% in the two- versus the three-verdict condition (see Table 2), respectively: $b = -0.74$, $OR = 0.48$, $z = -2.30, p = .022$, 95% CI [0.25, 0.89]; $\chi^2(1, N = 161) = 5.34, p = .021$. In Study 4, in the two-verdict condition, the conviction rate was 51.81% but dropped to 26.25% in the three-verdict condition (see Table 2). The statistical analysis showed that this difference was also statistically significant: $b = -1.11$, $OR = 0.33$, $z = -3.29$, $p < .001$, 95% CI [0.17, 0.63]; $\chi^2(1, N = 163) = 11.16$, $p < .001$.

Consider, finally, the analysis of Study 5's results (among laypeople). In the two-verdict condition, the conviction rate was 63.98% but dropped to 55.00% in the three-verdict condition (see Table 2 and Figure 1). As before, the statistical analysis indicated that this difference was statistically significant: $b = -0.37$, $OR = 0.69$, $z = -2.05$, $p = .041$, 95% CI [0.48, 0.98]. Similar findings were obtained under the chi-square test: $\chi^2(1, N = 501) = 4.19$, $p = .041$.

The N-Verdict Bias and Legal Expertise

Our central analysis investigated whether the decision-maker's legal expertise moderates the N-verdict bias. More specifically, we tested whether the N-verdict bias is less pronounced among legal professionals compared to nonprofessionals (i.e., law students and laypeople). For this purpose, we combined the data from Studies 1–5, resulting in 1,040 participants. We created two dummy-coded variables, with professionals as the control group. The analysis thus compared professional lawyers (Studies 1 and 2) to law students (Studies 3 and 4), and professional lawyers to laypeople (Study 5). We then built a logistic regression with "verdict" as the dependent variable, as in previous analyses. There were five independent variables: condition (two vs. three verdicts), the two dummy-coded variables, and the interaction between the dummy variables and the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

condition. These interactions between the populations and the experimental condition enabled us to test whether the experimental manipulation differed in magnitude across populations.

First, as in previous analyses, we found a main effect of "number of verdict-alternatives," such that participants were less inclined to convict in the three-verdict than in the two-verdict condition: $b = -0.87$, $OR = 0.42$, $z = -2.41$, $p = .016$, 95% CI [0.20, 0.84]. Second, we found that, compared to professionals, law students presented a higher conviction rate, $b = 1.35$, $OR = 3.85$, $z = 5.02$, $p < .001$, 95% CI [2.30, 6.60], and laypeople also tended to convict more than professionals, $b = 1.59$, $OR = 4.90$, $z = 6.30$, $p < .001$, 95% CI [3.02, 8.14]. Third, we found no significant interaction between population and condition. The results were not statistically significant either for the comparison between professional lawyers and law students ($b = -0.01$, $OR = 0.99$, $z = -0.03$, $p = .974$, 95% CI [0.43, 2.31]) or for the comparison between professional lawyers and laypeople ($b = 0.49$, $OR = 1.64$, $z = 1.22$, $p = .221$, 95% CI [0.75, 3.68]).

Finally, we repeated the above analysis, this time controlling for the two demographic variables we had collected: age and gender (coded 1 or 0 for female or male participants, respectively). This resulted in the inclusion of 977 participants (93.94% of the initial sample) who reported their age and their gender (i.e., as either a female or a male participant—a third category, "other," being selected by only three individuals). Broadly speaking, this analysis replicated the findings reported above: a significant effect of the number of verdict-alternatives and a difference between the populations in their conviction rates, but no interaction between the population and the experimental condition. First, we found a main effect of the number of verdict-alternatives, such that participants were less inclined to convict in the three-verdict than in the two-verdict condition: $b = -0.88$, $OR = 0.42$, $z = -2.20$, $p = .028$, 95% CI [0.18, 0.89]. Second, we found that, compared to professional lawyers, law students presented a higher conviction rate, $b = 1.05$, $OR = 2.86$, $z = 3.44$, $p < .001$, 95% CI [1.58, 5.26], and laypeople also tended to convict more than professionals, $b = 1.45$, $OR = 4.25$, $z = 5.48$, $p < .001$, 95% CI [2.56, 7.22]. Third, we found no significant interaction between population and condition either for the comparison between professional lawyers and law students ($b = -0.03$, $OR = 0.98$, $z = -0.06$, $p = .957$, 95% CI [0.40, 2.48]) or for the comparison between professional lawyers and laypeople ($b = 0.49$, $OR = 1.64$, $z = 1.12$, $p = .262$, 95% CI [0.70, 3.98]). Fourth, regarding the demographic variables, we found no significant effect by age: $b = -0.01$, $OR = 0.99$, $z = -1.15$, $p = .249$, 95% CI [0.98, 1.00]. We did, however, find a gender effect, whereby women convicted more often than men: $b = 0.31$, $OR = 1.36$, $z = 2.23$, $p = .026$, 95% CI [1.04, 1.78].

The Selection of Acquittal Verdicts

The formal analyses examined thus far all focused on assessing the N-verdict bias by comparing the conviction rate in the two- versus three-verdict condition, among each population. Interestingly, as shown in Table 2, not all ABD verdicts were rendered at the expense of convictions: Some were made at the expense of full acquittal. Comparing acquittals in the two-verdict condition with *full* acquittals in the three-verdict condition, our analyses showed that the rate of full acquittals was lower in the three-verdict condition than the rate of acquittals in the two-verdict condition. This finding was obtained for each population and each study, as assessed by both logistic regressions (ps $< .041$) and chi-square tests (ps $< .034$).

Discussion

Let us now examine the findings derived from Studies 1–5. We begin with the findings pertinent to our claims and then briefly mention a few that are outside the scope of our investigation. First, the presence of an intermediate verdict was appealing to a large portion of our participants, be they professional lawyers, law students, or laypeople. The fact that the (full) acquittal rate was lower in the three- than in the two-verdict condition (see Table 2) suggests that the selection of the ABD verdict was made, at least partly, at the expense of full acquittal. This pattern indicates that some participants who might have acquitted the defendant under the two-verdict condition opted for an intermediate verdict under the three-verdict condition. This behavior is permissible under the rational model exposed earlier, albeit not required by it.

Second, the ABD verdict was also selected at the expense of conviction verdicts. This claim is supported by the finding that the availability of an intermediate verdict (i.e., in the three-verdict condition) decreased the rate of conviction in comparison to the two-verdict condition. This comparison revealed a statistically significant effect within each population: professional lawyers (Studies 1 and 2), law students (Studies 3 and 4), and laypeople (Study 5). When each study was analyzed separately, the effects were statistically significant in Studies 3, 4, and 5 and in the expected direction, yet not significant in Studies 1 and 2 (likely due to a lack of statistical power).

Under the rational model, this effect constitutes a bias, since the decision to convict should not be affected by the number of acquittal verdict-alternatives available to the decision-maker but should be based solely on the available evidence and the required standard of proof. This pattern was evident among professional lawyers, law students, and laypeople alike, indicating the pervasive effect of bias within all the populations investigated.

Third, and most importantly, it appears that legal professionals are potentially just as swayed by the presence of an intermediate verdict as nonprofessionals. Neither of the two interactions we examined between the N-verdict bias and expertise proved statistically significant. In the first interaction, we compared the N-verdict bias between professional lawyers and law students. Recall that both groups read exactly the same materials, in the same language, in the same country. In the second interaction, we compared the N-verdict bias between professional lawyers and laypeople. Here, too, both groups read near-identical materials (with some adaptation, as detailed earlier, and in different languages, Hebrew and English). However, because these two groups came from different countries (Israel and the United States), this second comparison carried a confound between expertise and nationality that the first one did not, and we acknowledge this when interpreting the result. Our baseline assumption, though, was that legal expertise would reduce the extent of the bias regardless of nationality, and we found no evidence for the hypothesis that legal professionals are less impacted than nonprofessionals by the availability of an intermediate verdict. It therefore seems safe to conclude that legal expertise does *not* de-bias legal decision-makers.

Turning briefly to other findings that may be interesting—though they are outside the scope of our inquiry—first, we found that professional lawyers tended to acquit the defendant more than both law students and laypeople. Perhaps this tendency can be attributed

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

to their education, training, and experience, which may render them more capable of effectively scrutinizing the arguments and evidence presented to them by the prosecution, which bears the burden of proof in criminal trials. Be that as it may, these findings neither support nor undermine our claims about the N-verdict bias.

Second, it appears that the law students who participated in Study 3 exhibited a higher conviction rate than their counterparts in Study 4. We do not have a compelling explanation for this finding. One can only speculate that several variables may account for this difference, including situational factors such as the year of the undergraduate degree course they were studying or the day or hour when the experiment was run. The important finding, however, remains that availing of an intermediate verdict significantly decreased the conviction rate in both studies.

Finally, the conviction rates differed among women and men. Here, it is important to keep in mind that the individuals who participated in our studies may not represent the general population. Therefore, any future interpretation should be made with caution.

Study 6: The N-Verdict Bias and Instructions Regarding Reasonable Doubt

Regarding the purpose of Study 6, let us begin by emphasizing that the participants in the preceding five studies were deliberately given no standard of proof and no legal definition of murder: The materials neither used the phrase "beyond-reasonable-doubt" nor referred to any other standard of proof. As Studies 1–4 were designed to simulate decisions made by professional judges, the rationale for this intentional omission was that professional adjudicators and practicing lawyers decide without formal instructions or reminders of this kind. And, while Study 5 was conducted on laypeople from a different jurisdiction, we used identical materials to those of Studies 1–4 to enable comparison. The aim of Study 6, then, was to check whether the N-verdict bias persisted even when lay participants were given an explicit standard of proof and instructed that they should convict only if the charged offense was proven beyond-reasonable-doubt. In other words, Study 6 evaluated whether Study 5's lack of formal instructions about the standard of proof affected the generalizability of the findings of Studies 1–4 to lay decision-makers. It was preregistered at https://aspredicted.org/57x2-m4jd.pdf.

As for the legal definition of murder, our hypothetical case was carefully constructed so that the only fact in dispute between the prosecution and the defense was *identity*: whether the defendant was the person who committed the murder. The defense does not raise the claim of self-defense, and the evidence described in the studies does not support such a claim (there are no facts presented suggesting that the victim presented an imminent threat to the perpetrator, for instance, or that the perpetrator's use of force was necessary). Consequently, if a case like ours had reached a trial, then no jury, and certainly not professional judges, would have received any instructions on legal questions that were not raised by the parties and were irrelevant to the facts in dispute, such as self-defense.

Method

Participants

We recruited 428 U.S. lay participants (M_{age} = 42.32, *SD* = 12.38; majority, men: 50.23%), again via the Prolific platform. As in Study 5, participants were all American English-speaking individuals who completed a 12-min study in return for approximately $2.70. Of the initial 504 participants, 76 (15.08%) were excluded for failing comprehension checks, as per the preregistration specifications.

Procedure and Materials

The study was conducted online. Following their consent, participants were redirected to an external platform (Qualtrics), where they received specific instructions. Study 6 was identical to Study 5 in all respects except one: Before reading the case, participants were explicitly instructed that they "may find the defendant guilty only if the People have proven the charged offense beyond a reasonable doubt." (The Supplemental Materials attached to this article provide an exact image of the instruction screen.) Participants were asked to read the materials carefully and then render a verdict on the case, as if they were the judge, by selecting the relevant verdict on-screen to indicate their decision. Finally, they were required to answer two sociodemographic questions regarding their age and gender.

Although Study 6 supplied the standard of proof, participants were not given a definition of "reasonable doubt," for the term lacks a formal definition, and trial judges and counsel are explicitly discouraged from explaining its meaning to jurors. As early as 1880, the U.S. Supreme Court noted that "Attempts to explain the term 'reasonable doubt' do not usually result in making it any clearer to the minds of the jury" (*Miles v. United States*, 1880, p. 312). And the oft-cited case of *United States v. Glass* (1988) states,

> "Reasonable doubt" must speak for itself. Jurors know what is "reasonable" and are quite familiar with the meaning of "doubt." … It is, therefore, inappropriate for judges to give an instruction defining "reasonable doubt," and it is equally inappropriate for trial counsel to provide their own definition. (p. 387)

More recently, the U.S. Supreme Court emphasized that "the Constitution does not require that any particular form of words be used in advising the jury of the government's burden of proof" (*Victor v. Nebraska*, 1994, p. 5). In England and Wales, the official instructions to judges state, regarding the standard of proof: "If the jury ask for clarification of the standard, their question should be answered as shortly as possible" (Judicial College, 2026, pp. 5-1–5-2, point 6).

Regarding the difference between "not guilty" and "not proven" or between "full acquittal" and "acquittal for the benefit of the doubt," no specific instruction is encouraged in Israel, let alone required. And, in the official manual for juries in Scotland, judges were warned, "It is dangerous to attempt to explain any difference between the not proven and not guilty verdicts" (Judicial Institute for Scotland, 2024, pp. 121.1/133).

For similar reasons, we also limited ourselves to one specific set of verdicts and opted not to test other sets of possible alternative verdicts (such as an intermediate verdict with a different name or a third verdict that was not intermediate at all). Our consistent motivation was to stay as close as possible to the verdicts presented to actual legal decision-makers in real-world situations. Furthermore, to enable us to compare the results between the studies, it was essential to keep the set of verdicts constant.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Design

As in Studies 1–5, participants were randomly assigned to one of two conditions. In the *two-verdict condition*, they were asked to select either "to convict the defendant" or "to acquit the defendant." In the *three-verdict condition*, they were asked to select either "to convict the defendant," "to acquit the defendant on the benefit of the doubt," or "to fully acquit the defendant."

Results

Table 3 presents the percentages of participants selecting each type of verdict, by condition.

First, replicating the findings of Studies 1–5, analysis of the results of Study 6 showed that, when the opportunity to select an intermediate verdict was made available, a large portion of decision-makers did so. In the three-verdict condition, 48.87% of the verdicts were ABD.

Second, and more importantly, the analysis revealed that, as in Studies 1–5, the intermediate verdict was selected at the expense of some of the "guilty" verdicts, thereby reducing the overall conviction rate and reflecting the bias created by the number of verdict-alternatives. In the two-verdict condition, the conviction rate was 59.42% but dropped to 47.06% in the three-verdict condition. The logistic regression demonstrated that this difference was statistically significant: $b = -0.50$, $OR = 0.61$, $z = -2.55$, $p = .011$, 95% CI [0.41, 0.89]. This pattern was also revealed using a chi-squared test: $\chi^2(1, N = 428) = 6.56$, $p = .01$. Note that, in this study, as in Studies 1–5, if we compare acquittals in the two-verdict condition with *full* acquittals in the three-verdict condition, we can observe that the rate of full acquittal was lower in the latter. This finding was obtained both when analyzing the data with logistic regressions and with chi-square tests ($ps < .001$).

Finally, to explore whether the N-verdict bias was of a different magnitude when instructions were present (Study 6) rather than absent (Study 5), we combined the data of both studies in order to investigate the interaction between instructions (present vs. absent) and the experimental condition (two vs. three verdicts). This led to the inclusion of 929 participants. The study variable was dummy-coded (0 for Study 5 and 1 for Study 6), as was the condition variable (0 for the two-verdict condition and 1 for the three-verdict condition). We then built a logistic regression predicting participants' verdicts (coded 1 for conviction, 0 otherwise) from the study, the experimental condition, and their interaction.

The results indicated that, as in all previous analyses, the conviction rate was reduced in the three- versus the two-verdict condition: $b = -0.43$, $OR = 0.65$, $z = -3.23$, $p = .001$, 95% CI [0.50, 0.84]. The findings did not support the conclusion that participants convicted more often in one study than in the other: $b = -0.26$, $OR = 0.77$, $z = -1.91$, $p = .056$, 95% CI [0.60, 1.01]. Finally, and most importantly, no interaction was found between study and experimental condition: $b = -0.13$, $OR = 0.88$, $z = -0.47$, $p = .640$, 95% CI [0.52, 1.49].

Table 3
Results of Study 6: Distribution (%) of Verdicts by Condition

Condition	*N*	Conviction (%)	Acquittal (%)	ABD (%)
Two-verdict	207	59.42	40.58	
Three-verdict	221	47.06	4.07	48.87

Note. ABD = acquittal for the benefit of the doubt.

Discussion

The N-verdict bias was obtained even when participants were explicitly instructed that they could convict only in the absence of reasonable doubt. That is, this instruction did not neutralize the N-verdict bias. Just as in Studies 1–5 (where no such instructions were given), the conviction rate in Study 6 was lower when the intermediate verdict was available compared to when it was not. This study therefore demonstrates the pervasive effect of the bias. It also dispels the potential concern that the N-verdict bias documented in Studies 1–5 was due to the absence of an instruction about the required standard of proof.

Finally, we found no evidence in support of the possibility that the magnitude of the N-verdict bias is affected by the presence or absence of explicit instructions, as revealed by the nonsignificant interaction between study and condition. While combining the data of Studies 5 and 6 certainly does not enable a clean comparison (as the studies differed from each other—for instance, they were not run on the same date), if explicit instructions *had* been crucial for the production of the N-verdict bias, the bias would be of different magnitudes in the two studies.

Study 7: The Number of Verdicts and the Confidence Level Required for Conviction

The aim of Study 7 was to test the possibility that the mechanism underlying the N-verdict bias involves a change in the individual's internal threshold for conviction. Should this be the case, when the number of verdicts is increased from two to three by offering an intermediate option to "acquit on the benefit of the doubt," the internal degree of confidence that the individual deems necessary to convict will increase. And the level of confidence that people require for conviction will be higher under the three-verdict condition than under the two-verdict condition. We now turn to the methodology applied in Study 7, which was preregistered at https://aspredicted.org/pmdb-48tg.pdf.

Method

Participants

We recruited 466 U.S. lay participants ($M_{age} = 42.17$, $SD = 13.26$; majority, women: 51.29%) via Prolific. Participants were all American English-speaking individuals who completed a 2-min online study, in return for approximately $0.30. Of the initial 490 participants, 24 (4.90%) were excluded, per the preregistration specifications: 19 for failing comprehension checks and five for providing a zero value for the main dependent variable.

Procedure and Materials

In this online study, after giving their consent, participants were redirected to an external platform (Qualtrics), where they were given specific instructions. Unlike previous studies, Study 7 did not include any cases or rendering of verdicts. Instead, participants were asked to imagine they were a juror in a murder trial (in a country

where they had to choose between two verdicts or three, depending on the experimental condition to which they were randomly assigned). Like Study 6, they were told that jurors in the country in question should convict only if the people had proven beyond-reasonable-doubt that the defendant had committed the murder. To determine their internal threshold, they were then asked to give their opinion on how convinced they would need to be, as a juror, that the defendant committed the murder, in order to choose "Conviction," on a scale from 0 (*not convinced at all*) to 100 (*absolutely convinced*).

Design

Participants were randomly assigned to one of two conditions. In the *two-verdict condition*, they were told that, in the country in question, the jury could choose "to convict the defendant" or "to acquit the defendant." In the *three-verdict condition*, they were told that the jury could choose "to convict the defendant," "to acquit the defendant on the benefit of the doubt," or "to fully acquit the defendant." In both conditions, participants were asked to indicate the level of confidence they would need in order to convict.

Results

The dependent variable was "internal confidence level," measured on a scale ranging from 0 to 100. First, across conditions, the mean threshold was high ($M = 93.72$, $SD = 13.15$), suggesting that participants generally required a high level of confidence in order to render a conviction.

However, the mean required confidence did not significantly differ between the two conditions: 94.30 and 93.14 in the two- and the three-verdict conditions, respectively. An independent-samples t test revealed that this difference was not statistically significant: $t(464) = 0.95$, $d = 0.09$, $p = .343$.

Discussion

If the N-verdict bias were due to an increase in the confidence required for conviction, the number of verdict-alternatives would affect the degree of confidence that decision-makers need to obtain to convict a defendant. The results of Study 7 do not support the existence of such a relationship, as the mean required level of certainty was *no* higher when the "acquittal on the benefit of the doubt" verdict was made available.

This conclusion should nevertheless be interpreted with caution. Unlike Studies 1–6, Study 7 did not require participants to render a verdict in response to a specific criminal case. Rather, they were asked to report, in principle, the level of confidence they believed they would need in order to convict a defendant. Thus, Study 7 necessarily assumed that people can meaningfully report an abstract confidence-threshold and that this threshold is related to the one they would apply when evaluating concrete evidence.

Note, however, that this design also offered an important advantage: Had participants reported their confidence-threshold after deciding a specific case, their responses might have been adjusted post hoc to fit the verdict they had just selected. Posing the question in the abstract therefore enabled us to test whether the mere availability of an intermediate verdict, independent of any details about a specific case, induced a different general mindset regarding the confidence required for conviction, compared to a two-verdict scenario.

General Discussion

Rational decision making dictates that the probability of conviction should not be affected by the number of verdict-alternatives available to legal decision-makers. To determine whether this requirement is honored in practice, the present research tested the hypothesis that the number of verdict-alternatives does, indeed, change the probability of conviction—a phenomenon we refer to as the N-verdict bias.

Central to our research was the role of legal expertise in the N-verdict bias. In jurisdictions where verdicts are rendered by professional judges rather than lay jurors, the preference for professional decision-makers is often justified on the premise that they will have the capacity to adhere strictly to the rational model of legal decision making and render a verdict of conviction under the sole circumstance that the evidence is sufficient to prove the defendant's guilt beyond-reasonable-doubt. The unspoken corollary to this is that lay juries may be susceptible to potential bias in their verdicts, unlike professional judges who ostensibly are not.

This being the case, our two central research questions were (a) whether the N-verdict bias affects not only lay jurors but also professional decision-makers and (b), more specifically, whether it is attenuated by the individual's legal expertise. We also tested two related claims: that instructing decision-makers that they may find the defendant guilty only if the people have proven the charged offense beyond-reasonable-doubt de-biases their decisions and that the N-verdict bias can be explained by an increase in decision-makers' internal threshold for conviction, which is triggered when the number of verdict-alternatives is increased from two to three.

Our studies, involving participants with different levels of legal decision-making expertise, revealed that an increase in the number of verdict-alternatives affected not only laypeople's decision making but also that of law students and practicing legal professionals. Importantly, we did not find support for the hypothesis that legal expertise moderates the effect of N-verdict bias, as no variation in the strength of the effect was detected when comparing the effect among professionals to that of law students and laypeople. This suggests that N-verdict bias is likely to be at work also among professional judges. This finding is also consistent with those of Rachlinski and Wistrich (2018), who showed that professional judges are not immune to the biases that affect laypeople.

In addition, our results do not support the claim that emphasizing the "beyond-reasonable-doubt" requirement (by including instructions) helps to de-bias decision-makers. In Study 6, the N-verdict bias was obtained even when participants were explicitly instructed to take this issue into account. Finally, our results do not support the claim that an intermediate verdict might affect the conviction rate by raising decision-makers' internal confidence-threshold vis-à-vis the decision to convict.

Constraints on Generalizability

Our studies present certain limitations that need to be acknowledged. First, the participants were asked to select between the following outcomes: "convict," "acquit," and "ABD." These categories differ from those usually considered in the existing research ("guilty,"

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

"not guilty," and "not proven"), and therefore, the findings from our study might not be applicable to jurisdictions that rely on other categories. However, we expect our main finding—that the availability of an intermediate verdict influences legal professionals' conviction rate in the Israeli system—to apply also to such professionals in other jurisdictions. This is because, irrespective of what their verdicts happen to be called, all criminal trials are designed to answer the same legal question: whether the available evidence suffices to convict the defendant. The generalizability of our findings is also supported by the various studies that have been conducted on laypeople that found a similar effect with the "not proven" verdict (Curley et al., 2019; Curley, Munro, et al., 2022; Hope et al., 2008; Ormston et al., 2019). Merely changing the verdict's names is, hence, unlikely to nullify the effect of the intermediate verdict on the conviction rate.

Second, our studies did not cover other legal mechanisms that could be considered alternative verdicts, such as convictions for lesser offenses or jury nullification. Vidmar (1972), for example, showed that providing jurors with the option to convict the defendant of a lesser offense substantially reduced acquittal rates. However, from a legal perspective, a conviction of a lesser offense is not an *alternative verdict* of conviction—it is the same verdict of conviction, albeit relating to a different offense. Both convictions are identical in all procedural aspects, particularly the need for proof beyond-reasonable-doubt to secure them, and the difference between them is attributable to substantive criminal law rather than to criminal procedure. That said, psychologically, a lesser offense may engage mechanisms much like those discussed in the introduction, and we leave this question to future studies. Jury nullification is another matter entirely. This is where a jury finds the evidence sufficient to prove the defendant's guilt beyond-reasonable-doubt but opts to acquit nonetheless (for instance, because they think the law is unfair, the punishment too harsh, and so on). Jury nullification is a complicated issue, and unlike "not proven," it is unclear whether jurors must be informed of its existence; hence, it needs to be analyzed separately.

Third, most of our studies involved the same hypothetical scenario, so the effect that we measured might not occur (or might occur to a different extent) when different facts and offenses apply. However, it was essential to use the same materials, in order to make a valid comparison of conviction rates and the N-verdict bias among participants with different levels of expertise. Notwithstanding, our findings are not limited to the scenario we used in our studies: They are applicable to other factual scenarios and criminal offenses. In making this claim, we draw support from the abovementioned previous studies that dealt with the N-verdict bias and demonstrated the effect of the "not proven" outcome on conviction rates in various scenarios and for different offenses (such as assault or rape). We do not claim, however, that the facts and offenses involved have *no* impact on the strength of the N-verdict bias. On the contrary, in future research, we aim to identify precisely how these factors may interact with the N-verdict bias and to explain why. For now, the important point is that, whatever these interactions are, there is no evidence to support possible concerns that our findings are limited solely to the particular hypothetical scenario that we used in our studies.

Fourth, our studies did not fully explore how the N-verdict bias might be affected by the existence, content, or location of various instructions about either (a) the definition of "ABD," (b) the meaning of "beyond-reasonable-doubt," or (c) the structure of the offense of which the accused is charged. On the one hand, the content of instructions and definitions provided at trial, ostensibly to assist decision-makers, is highly disputed, and some scholars question whether they should be provided at all. On the other hand, as already explained, our rationale for conducting our studies *without* such instructions was that courts are generally advised against providing them. This issue remains contested, and reforming the system to provide well-structured instructions could arguably minimize the N-verdict bias (or perhaps even eliminate it). Study 6 constituted a first step toward resolving this question by testing this very argument. It showed that the N-verdict bias cannot be dismissed merely as attributable to the lack of instructions. All the same, perhaps instructions can still be helpful, but only if they are carefully constructed. Horowitz and Kirkpatrick (1996), for example, demonstrated that providing an accessible and understandable definition of the "beyond-reasonable-doubt" standard, such as "firmly convinced," assisted the jury in identifying and applying the correct legal standard. Before any recommendation can be made, further research is therefore needed to determine whether, and how, instructions may affect the N-verdict bias.

Fifth, our studies were limited to the decision of individual jurors—they did not examine a jury's *collective* decision making. However, previous studies found that the availability of "not proven" rendered individual jurors significantly less likely to convict, both before and after deliberation of the jury as a group (see the aforementioned Ormston et al., 2019). At any rate, group-based jury decisions are subject to jurisdiction-specific rules on jury size (usually 12 in the United States and England, but 15 in Belgium and Scotland, for instance), the majority required (unanimity in the United States, supermajority in England, and simple majority in Austria and Brazil), and the level of legal expertise (jury members in the United States, England, and Scotland are all laypeople, while three members of the jury in Belgium must be professional judges). Moreover, group decisions are influenced by various psychological processes related to the intra-personal nature of the process that are outside the scope of our investigation and merit extensive future research. For some interesting findings on the interaction between the number of verdicts (two vs. three) and the type of majority required (simple vs. unanimity), see Ormston et al. (2019).

Last, although our findings are intended to provide insights relevant to jurisdictions in which criminal verdicts are rendered by professional judges, we were prevented from conducting any of our studies on such professionals. The judiciary under study, that of Israel, is known for its policy of noncooperation with empirical studies such as ours. While regrettable, it is important to emphasize that the N-verdict bias that was measured among legal adjudicators and practicing lawyers, as well as law students and laypeople, was of a similar magnitude across *all* these subsamples. This finding suggests that the N-verdict bias is not moderated by legal expertise, thereby providing substantial support for our claim that professional judges might also be subject to it. Previous research offers no evidence relevant to this claim, since it has never addressed professional legal decision making, thus rendering our findings a tangible empirical advancement toward better understanding this phenomenon, vis-à-vis the extant literature. In sum, while these limitations certainly restrict the generalizability of our findings, they do not undermine their theoretical or practical implications.

Should Intermediate Verdicts Be Encouraged or Avoided?

While the impact of intermediate verdicts on the conviction rate undoubtedly constitutes a bias, this finding is insufficient, on its

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

own, to conclude that such verdicts should be avoided. To reach such a conclusion, it would be necessary to assess the impact of intermediate verdicts on three different defendant populations, which we describe next.

The first population consists of (in fact) guilty defendants who were mistakenly acquitted on the benefit of the doubt in the three-verdict system but would have been (correctly) convicted in a two-verdict system. In this case, the availability of an intermediate verdict helps guilty defendants evade the law. The second population consists of (in fact) innocent defendants who were acquitted on the benefit of the doubt in the three-verdict system but would have been fully acquitted in a two-verdict system. In such cases, rendering an intermediate verdict has unjust adverse effects on the innocent defendant: While it carries no criminal sanction, it can still have some serious ramifications for the individual. In the recently abolished Scots system, it affected the good name and reputation of the acquitted innocent, undermining their social standing and employment prospects. In Israel, the intermediate verdict has additional formal consequences, since it denies the acquitted innocent compensation for wrongful prosecution and can be used against them in child custody disputes and permanent residency applications (Rabin & Vaki, 2021). These ramifications should be given considerable weight since the results of our studies showed that the presence of an intermediate verdict reduces not only the conviction rate but also the probability of *full* acquittal, thereby significantly reducing the chances of an innocent defendant being unequivocally acquitted, as they may have been in the two-verdict system.

For both the former and latter populations, the availability of an intermediate verdict leads to socially undesirable consequences. By contrast, for a third population, the outcomes of intermediate verdicts are markedly desirable. The third population consists of (in fact) innocent defendants who were acquitted (usually on the benefit of the doubt) in the three-verdict system but may have been falsely convicted under a two-verdict system. We contend that the effect of the availability of an intermediate verdict on this third population is critical when it comes to making informed recommendations to policymakers on whether such verdicts should be avoided or encouraged.

Conclusion

All criminal trials are built on the same premise: Whether or not a defendant is convicted should turn solely on the strength of the evidence, not on the menu of verdicts the decision-maker happens to be offered. Our findings show that this premise does not hold in practice. Across studies spanning the full range of legal expertise (laypeople, law students, practicing lawyers, and Bar disciplinary-panel adjudicators), simply making an intermediate verdict available lowered the rate of conviction. Translated into real-world terms, this means that some defendants who would be convicted when the only options are to convict or to acquit would *not* be convicted when a third, intermediate verdict is added to the slate. Throughout our empirical studies, because the evidence and the charge were held constant, the impact of the number of verdict-alternatives on conviction rates can be understood to reflect a genuine bias—a systematic departure from the very standard that the law prescribes. Furthermore, the bias withstood an explicit instruction to lay participants to convict only when guilt was proven beyond-reasonable-doubt (Study 6), and it did not seem to operate by raising the confidence level that lay participants said they would need before convicting (Study 7).

What gives our findings their weight is the fact that legal expertise offered no protection against the effect of the number of verdicts available: Even among seasoned professionals, the conviction rate still fell from approximately 27% to 13%. This reaffirmed our suspicion that the effect is not confined to untrained jurors. It appears instead to be a feature of legal decision making itself, one that probably extends to professional judges and that bears on every jurisdiction in which trained professionals decide whether to prosecute or convict, even where laypeople are also involved as legal decision-makers. Our results join Rachlinski and Wistrich's (2018) finding that experienced judges share many of the biases that affect ordinary jurors, and lend support to the position that legal professionals do *not* make more rational decisions than nonprofessionals.

None of this settles whether intermediate verdicts should be retained or abolished. That a reform shifts conviction rates does not, on its own, reveal whether the shift is for better or worse; the answer turns on whom the added option shields and whom it harms, and above all on those innocent defendants who are spared a wrongful conviction because an intermediate verdict was available. It is this latter group that we plan to investigate next.

References

Barbato, J. M. (2005). Scotland's bastard verdict: Intermediacy and the unique three-verdict system. *Indiana International & Comparative Law Review*, *15*(3), 543–582. https://doi.org/10.18060/17848

Barneron, M., & Yaniv, I. (2020). Advice-giving under conflict of interest: Context enhances self-serving behavior. *Journal of Experimental Social Psychology*, *91*, Article 104046. https://doi.org/10.1016/j.jesp.2020.104046

Bray, S. (2005). Not proven: Introducing a third verdict. *University of Chicago Law Review*, *72*(4), 1299–1329. https://chicagounbound.uchicago.edu/uclrev/vol72/iss4/4/

Chalmers, J., Leverick, F., & Munro, V. E. (2021). A modern history of the not proven verdict. *Edinburgh Law Review*, *25*(2), 151–172. https://doi.org/10.3366/elr.2021.0692

Curley, L. J., MacLean, R., Murray, J., Laybourn, P., & Brown, D. (2019). The bastard verdict and its influence on jurors. *Medicine, Science and the Law*, *59*(1), 26–35. https://doi.org/10.1177/0025802418811740

Curley, L. J., Munro, J., Frumkin, L. A., & Turner, J. (2021). Informing reform: The views of legal professionals on the unique aspects of Scottish law. *Medicine, Science and the Law*, *61*(4), 256–265. https://doi.org/10.1177/0025802421992913

Curley, L. J., Munro, J., Turner, J., Frumkin, L. A., Jackson, E., & Lages, M. (2022). Proven and not proven: A potential alternative to the current Scottish verdict system. *Behavioral Sciences & the Law*, *40*(3), 452–466. https://doi.org/10.1002/bsl.2568

Curley, L. J., Murray, J., MacLean, R., Munro, J., Lages, M., Frumkin, L. A., Laybourn, P., & Brown, D. (2022). Verdict spotting: Investigating the effects of juror bias, evidence anchors and verdict system in jurors. *Psychiatry, Psychology and Law*, *29*(3), 323–344. https://doi.org/10.1080/13218719.2021.1904450

Guthrie, C. (2002). Panacea or Pandora's box? The costs of options in negotiation. *Iowa Law Review*, *88*(3), 601–654. https://doi.org/10.2139/ssrn.358406

Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2001). Inside the judicial mind. *Cornell Law Review*, *86*(4), 777–830. https://scholarship.law.cornell.edu/facpub/814/

Hope, L., Greene, E., Memon, A., Gavisk, M., & Houston, K. (2008). A third verdict option: Exploring the impact of the not proven verdict on mock

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

juror decision making. *Law and Human Behavior*, *32*(3), 241–252. https://doi.org/10.1007/s10979-007-9106-8

Horowitz, I., & Kirkpatrick, L. (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*(6), 655–670. https://doi.org/10.1007/BF01499236

Jackson, E., Curley, L., Leverick, F., & Lages, M. (2024). The effect of verdict system on juror decisions: A quantitative meta-analysis. *Psychiatry, Psychology and Law*, *32*(1), 23–41. https://doi.org/10.1080/13218719.2023.2272912

Judicial College. (2026). *The Crown Court compendium—Part I: Jury and trial management and summing up*. https://www.judiciary.uk/wp-content/uploads/2022/06/Crown-Court-Compendium-Part-I-Oct-25-Mar-26-update.pdf

Judicial Council of the United States Eleventh Judicial Circuit. (2025). *Court's instructions to the jury*. https://www.ca11.uscourts.gov/sites/default/files/courtdocs/clk/FormCriminalPatternJuryInstructionsRevisedSEP2025.pdf

Judicial Institute for Scotland. (2024). *e-Jury manual*. https://judiciary.scot/docs/librariesprovider3/judiciarydocuments/judicial-institute-publications/jury-manual-pdf-version-3-september-2024.pdf

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Miles v. United States, 103 U.S. 304 (1880). https://supreme.justia.com/cases/federal/us/103/304/

Ormston, R., Chalmers, J., Leverick, F., Munro, V., & Murray, L. (2019). *Scottish jury research: Findings from a large-scale mock jury study*. Scottish Government. https://tinyurl.com/4buz5e55

Phalen, H. (2018). Overcoming the opposition to a third verdict: A call for future research on alternative acquittals. *Arizona State Law Journal*, *50*, 401–430. https://arizonastatelawjournal.org/wp-content/uploads/2018/05/Phalen_Pub.pdf

Picinali, F. (2022). *Justice in-between: A study of intermediate criminal verdicts*. Oxford University Press. https://doi.org/10.1093/oso/9780198864592.001.0001

Rabin, Y., & Vaki, Y. (2021). Two kinds of acquittals—Different kinds of doubts. *Criminal Law Forum*, *32*(1), 97–123. https://doi.org/10.1007/s10609-020-09407-9

Rabin, Y., Vaki, Y., & Rahav, G. (2025). זיכוי מחמת הספק– האם הגיעה העת לביטולו? [Acquittal for the benefit of the doubt: Has the time come for abolition?] *Haifa Law Review*, *18*, 285–341. https://law.haifa.ac.il/wp-content/uploads/2025/09/din18-05-vaki.pdf

Rachlinski, J. J., & Wistrich, A. J. (2018). Gains, losses, and judges: Framing and the judiciary. *Notre Dame Law Review*, *94*(2), 521–582. https://scholarship.law.nd.edu/ndlr/vol94/iss2/2/

Schauer, F., & Spellman, B. A. (2017). Analogy, expertise, and experience. *University of Chicago Law Review*, *84*(1), 249–268. https://chicagounbound.uchicago.edu/uclrev/vol84/iss1/12

Scottish Government. (2021). *The not proven verdict and related reforms: Consultation*. https://www.gov.scot/publications/not-proven-verdict-related-reforms-consultation/

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, *53*(2), 252–266. https://doi.org/10.1016/0749-5978(92)90064-E

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*(2), 158–174. https://doi.org/10.1086/209205

Smithson, M., Deady, S., & Gracik, L. (2007). Guilty, not guilty, or …? Multiple options in jury verdict choices. *Journal of Behavioral Decision Making*, *20*(5), 481–498. https://doi.org/10.1002/bdm.572

Teichman, D., Zamir, E., Nadler, J., Katz, G., & Tor, A. (2023). Biases in legal decision-making: Comparing prosecutors, defense attorneys, law students, and laypersons. *Journal of Empirical Legal Studies*, *20*(3), 543–584. https://doi.org/10.1111/jels.12365

Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press.

United States v. Glass, 846 F.2d 386 (7th Cir. 1988). https://law.justia.com/cases/federal/appellate-courts/F2/846/386/396745/

Victims, Witnesses, and Justice Reform (Scotland) Act 2025, asp 12. (2025). https://www.legislation.gov.uk/asp/2025/12

Victor v. Nebraska, 511 U.S. 1 (1994). https://supreme.justia.com/cases/federal/us/511/1/

Vidmar, N. (1972). Effects of decision alternatives on the verdicts and perceptions of simulated jurors. *Journal of Personality and Social Psychology*, *22*(2), 211–218. https://doi.org/10.1037/h0032605

Received October 19, 2025
Revision received July 24, 2026
Accepted July 30, 2026 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.