REACHING A VERDICT:
EMPIRICAL EVIDENCE OF THE
CRUMBLING CONVENTIONAL WISDOM
ON CRIMINAL VERDICT FORMAT

AVANI MEHTA SOOD

ONLINE APPENDIX

The stakeholder survey was administered online, using the Qualtrics platform.[277] Survey respondents on average completed 86% of the survey, and the majority of participants (78%) completed the entire survey. The sample sizes for the reported analyses vary according to the number of respondents who provided data on each measure.

Statistical Terminology

The reported results highlight key significant findings. A result is *statistically significant* when it "is very unlikely to have occurred merely by chance."[278] In other words, statistically significant effects reflect "reliable and detected" differences that "allow[] us to assume that what happened in one sample of people is likely to hold true in other samples of people."[279] Statistical significance is measured by the probability value of *p*, for which the "traditional criterion" in psychological sciences is $p < .05$.[280] This indicates that there is a less than 5% probability that the reported effects are due to chance. Results with *p* values between .05 and .10 are considered marginally significant.[281]

While statistical significance indicates that observed differences are "larger than expected by chance," *effect sizes* are reported to "provide an indication of *how large* the effect actually is."[282] Effect sizes can be "useful for determining the practical or theoretical importance of an effect, the relative contribution of different factors or the same factor in different

---

[277] *See* QUALTRICS, https://www.qualtrics.com/education [https://perma.cc/KC2V-7N72].
[278] FREDERICK J. GRAVETTER & LARRY B. WALLNAU, STATISTICS FOR THE BEHAVIORAL SCIENCES 211, 241 (10th ed. 2017).
[279] DANA S. DUNN, RESEARCH METHODS FOR SOCIAL PSYCHOLOGY 281 (2d ed., Wiley 2013).
[280] GRAVETTER & WALLNAU, *supra* note 278, at 242.
[281] *See id.* at 241–42.
[282] *Id.* at 388 (emphasis added).

circumstances, and the power of an analysis."[283]

Error bars in the graphs represent the *standard error* of the means, which provides "a method for defining and measuring sampling error."[284] In other words, the error bars indicate the extent to which the *sample mean* (e.g., the average response of the surveyed stakeholders) may differ from the unknown *population mean* (e.g., the average response of the stakeholder population at large).[285] As the size of a sample increases, the sample mean tends to better approximate the population mean, so the standard error becomes smaller.[286]

Statistical Tests

Differences between two discrete groups on scale measures are examined using *two-tailed* t-*tests*, with independent sample *t*-tests for between-group comparisons and paired sample *t*-tests for within-group comparisons.[287] Effect sizes for *t*-tests are reported using Cohen's *d*.[288]

Differences between and within more than two groups on scale measures are examined using *analysis of variance* (ANOVA).[289] Levene's test is employed to confirm homogeneity of variance—that the variance of dependent variables is equal across the groups being compared.[290] If Levene's testing indicates a violation of the homogeneity assumption, the Welch test is used to accommodate unequal variances.[291] Post-hoc tests are conducted after ANOVAs, "to determine exactly which mean differences are significant and which are not."[292] The Tukey HSD post-hoc method[293] is employed when homogeneity of variance can be assumed, and the Games-

---

[283]   Catherine O. Fritz & Jennifer J. Richler, *Effect Size Estimates: Current Use, Calculations, and Interpretation*, 141 J. EXPERIMENTAL PSYCH.: GEN. 2, 2 (2012).

[284]   GRAVETTER & WALLNAU, *supra* note 278, at 211.

[285]   *Id.* at 282.

[286]   *Id.* at 212.

[287]   *Id.* at 302–09, 337–42 (explaining *t*-tests).

[288]   General interpretation guidelines for Cohen's *d* suggest that .20 indicates a small effect, .50 indicates a medium effect, and .80 or above indicates a large effect. Jacob Cohen, *A Power Primer*, 112 PSYCH. BULL. 155, 157 (1992); GRAVETTER & WALLNAU, *supra* note 278, at 253, 316–17.

[289]   GRAVETTER & WALLNAU, *supra* note 278, at 366–74, 415–19 (explaining ANOVA and repeated-measures ANOVA).

[290]   GEOFFREY KEPPEL & THOMAS D. WICKENS, DESIGN AND ANALYSIS: A RESEARCHER'S HANDBOOK 147–51 (4th ed. 2004); *see generally* H. Levene, *Robust Tests for Equality of Variance*, *in* I. Olkin (ed.), CONTRIBUTIONS TO PROBABILITY AND STATISTICS: ESSAYS IN HONOR OF HAROLD HOTELLING (I. Olkin ed., Stan. Univ. Press 1960).

[291]   *See* KEPPEL & WICKENS, *supra* note 290, at 155; *see generally* B.L. Welch, *The Significance of the Difference Between Two Means When the Population Variances Are Unequal*, 29 BIOMETRIKA 350–61 (1938).

[292]   GRAVETTER & WALLNAU, *supra* note 278, at 393.

[293]   *Id.* at 394.

Howell post-hoc method is employed when homogeneity of variance cannot be assumed.[294] Effect sizes for ANOVAs are reported using eta-squared (η^2).[295]

Differences between and within groups on categorical measures are examined using *chi-square tests*, with Pearson chi-square for independent groups and McNemar-Bowker chi-square for paired groups.[296] Effect sizes for chi-square tests are reported using Cramer's V.[297] Adjusted standardized residuals (ASRs) are reported for chi-square comparisons to indicate the strength of significant differences between values.[298]

The degree and direction of relationships between scaled variables are examined with *Pearson correlation* and *linear regression*.[299] Finally, *mediation analysis*—which examines whether the relationship between two variables (the independent and dependent variables) is mediated by another variable (the mediator variable)—is conducted using the Sobel method. This approach tests the difference between the total effect (the independent-dependent variable relationship) and the direct effect (the independent-dependent variable relationship "after 'controlling for' the mediator").[300]

---

[294] *See* P.A. Games, H.J. Keselman & J.C. Rogan, *A Review of Simultaneous Pairwise Comparisons*, 37 STATISTICA NEERLANDICA 54–55 (1983); *see generally* P.A. Games & J.F. Howell, *Pairwise Multiple Comparison Procedures with Unequal N's and/or Variances: A Monte Carlo Study*, 1 J. EDUC. STATS. 113 (1976).

[295] General interpretation guidelines for η^2 suggest that .01 indicates a small effect, .06 indicates a medium effect, and .14 or above indicates a large effect. *See* GRAVETTER & WALLNAU, *supra* note 278, at 389; Fritz & Richler, *supra* note 283, at 10.

[296] GRAVETTER & WALLNAU, *supra* note 278, at 561–66, 573–78 (explaining chi-square).

[297] General interpretation guidelines for Cramer's V depend on the degrees of freedom in the analysis, with smaller V values counting for more as degrees of freedom increase. For example, if there is one degree of freedom, a Cramer's V of .10 indicates a small effect, .30 indicates a medium effect, and .50 indicates a large effect; but if there are five degrees of freedom, .05 indicates a small effect, .13 indicates a medium effect, and .22 indicates a large effect.

[298] An ASR of +/- 1.9 indicates a statistically significant difference; an ASR of +/- 3.0 suggests a large significant difference.

[299] BARBARA G. TABACHNICK & LINDA S. FIDELL, USING MULTIVARIATE STATISTICS 17–18 (5th ed. 2007); GRAVETTER & WALLNAU, *supra* note 278, at 489–90, 533–34 (explaining Pearson correlation and regression). General interpretation guidelines suggest that a Pearson's *r* of between +/- .1 and .3 indicates a small effect, .3 and .5 indicates a medium effect, and .5 or above indicates a large effect.

[300] TABACHNICK & FIDELL, *supra* note 299, at 159–61; *see generally* Michael E. Sobel, *Asymptotic Intervals for Indirect Effects in Structural Equation Models*, *in* SOCIOLOGICAL METHODOLOGY 290–312 (Samuel Leinhardt ed., 1982) (introducing and describing Sobel's methodology); *Calculation for the Sobel Test*, https://quantpsy.org/sobel/sobel.htm [https://perma.cc/D8PG-3NCL] ("Mediation . . . occur[s] when (1) the IV [independent variable] significantly affects the mediator, (2) the IV significantly affects the DV [dependent variable] in the absence of the mediator, (3) the mediator has a significant unique effect on the DV, and (4) the effect of the IV on the DV shrinks upon the addition of the mediator to the model.").

Abbreviations

The following abbreviations are used to describe variables in the statistical analyses:

Verdict Format Views

- **SV-Crim** = View on using special verdicts in criminal cases
- **GV-Crim** = View on using general verdicts in criminal cases
- **SV-Civ** = View on using special verdicts in civil cases

Stakeholders

- **Prosec** = Prosecutors
- **Judge** = Trial and appellate judges combined
  - **TriJ** = Trial judges
  - **AppJ** = Appellate judges
  - **J_Pros** = Judges with former prosecution experience
  - **J_CrimD** = Judges with former criminal defense experience
  - **J_Civ** = Judges with former civil litigation experience
- **CrimProf** = Criminal law professors
- **CivLit** = Civil litigators
- **CivProf** = Civil law professors
- **Lay** = Jury-eligible lay citizens
- **CrimSci** = Criminal science experts
- **CrimD** = Criminal defense attorneys (private and public combined)
  - **CrimD-Priv** = Private criminal defense attorneys
  - **CrimD-Pub** = Public defenders
- **LawStu** = Law students
- **CrimParty** = Party personally involved in a criminal case
  - **CrimDef** = Former criminal defendants
  - **CrimV** = Former victims in criminal cases
- **Jur** = Former jurors
  - **CrimJur** = Former criminal jurors
  - **CivJur** = Former civil jurors

Group Combinations

- **Stakeholder** = the ten primary stakeholder groups (presented in Table 1 of the Article)

- **CrimTri** = repeat legal actors in criminal trials: trial judges, prosecutors, and criminal defense attorneys (private and public)
- **CrimLaw** = criminal law professionals: judges (trial and appellate), prosecutors, criminal defense attorneys (private and public), and criminal law professors
- **Litig** = all litigators: prosecutors, criminal defense attorneys (private and public), and civil litigators (plaintiff and defense)

***

[i] Age: $M = 50.41$, $SD = 14.05$; Mode = 45 (25% of the respondents were under 40, 45% were between 40–59, and 30% were 60 or older).

[ii] Geographic region: 43% West, 23.5% South, 17.5% Midwest, 16% Northeast.

[iii] Less than 0.5% of the survey respondents (a total of four individuals) self-identified as non-binary.

[iv] The respondents of color were 4% each Black, Asian/Middle Eastern-North African, and Hispanic/Latin, with the remaining 3% of mixed racial/ethnic identities.

[v] Ideology: $M = 2.92$, $SD = 1.47$. Mean political ideologies differed by respondents' gender, geographic region, and stakeholder group as follows: Female respondents were significantly more liberal than male respondents. Ideology by Gender: $F(1, 1084.41) = 44.60$, $p < .001$, $\eta^2 = .03$. Respondents from the Northeast and West were significantly more liberal than those from the South and Midwest. Ideology by Region: $F(3, 502.74) = 12.94$, $p < .001$, $\eta^2 = .03$, Games-Howell: Northeast vs. Midwest at $p < .001$, Northeast vs. South at $p = .03$, West vs. Midwest & South at $p < .001$. Prosecutors were the most politically conservative stakeholder group, but their mean ideology score did not differ significantly from that of judges. Ideology by Stakeholder: $F(9, 389.92) = 25.73$, $p < .001$, $\eta^2 = .15$, Games-Howell: Prosec vs. other groups, except vs. Judge at $p < .001$. Criminal law professors and public defenders were the most politically liberal groups, and their mean ideology scores differed significantly from those of prosecutors and judges ($p < .001$), private criminal defense attorneys ($p = .006$ vs. CrimProf, $p = .01$ vs. CrimD-Pub), and civil litigators ($p = .01$ vs. CrimProf, $p = .02$ vs. CrimD-Pub).

[vi] Legal professionals' year of law school graduation: $M = 1994$, SD: 13.62; Mdn. = 1995; Mode = 1985. Legal professionals' age: Range = 24 to 87; $M = 52.67$, SD: 12.91; Mdn. = 52; Mode = 65.

[vii] Legal practitioners' area of practice: 76% criminal, 11% civil, 13% both.

[viii] Legal practitioners' level of practice: 80% trial, 6% appellate, 15% both.

[ix] Legal practitioners' jurisdiction of practice: 75% state, 12% federal, 13% both.

[x] Ideology of CrimD-Priv vs. CrimD-Pub: $F(1, 359) = 13.11$, $p < .001$, $\eta^2 = .04$.

[xi] Litigation experience of CrimD-Priv vs. CrimD-Pub: $X^2(3, 473) = 27.25$, $p < .001$, $V = .24$; CrimD-Priv ASR = 5.0 for CrimD-Civ experience, -3.7 for only CrimD experience.

[xii] The majority of the criminal science experts (73%) had a PhD or multiple advanced degrees in fields relating to science and law (such as a PhD or PsyD combined with a JD, MA, and/or MSc); 15% had only a MA or MSc; 12% had only a JD.

[xiii] Respondents who identified as being both former defendants and victims in criminal cases were excluded from the analyses, which sought to compare defendant-victim views.

[xiv] Stakeholders' views on verdict format are described as *supportive* or *opposed* when the 95% confidence intervals of their means fall above or below the neutral midpoint of the 5-point scale, respectively. Mean views are described as *neutral* when their 95% confidence intervals overlap with the neutral midpoint of the scale.

[xv] Correlation between SV-Crim & GV-Crim: $r(1536) = -.66$, $p < .001$.

[xvi] SV-Crim vs. GV-Crim: $X^2(10, N = 1538) = 97.46$, $p < .001$, $V = .47$.

[xvii] SV-Civ: $M = 3.69$, $SD = 1.00$. Respondents expressed more mean support for special civil verdicts than they did for either general or special criminal verdicts. SV-Civ vs. SV-Crim: $t(1020) = 7.65$, $p < .001$, Cohen's $d = .24$; SV-Civ vs. GV-Crim: $t(1020) = 6.95$, $p < .001$, Cohen's $d = .22$. Among the ten stakeholder groups, prosecutors expressed the least mean support for the use of special verdicts in civil cases ($M = 3.21$, $SD = 1.04$), and significantly less so than all other groups except jury-eligible lay citizens and law students. SV-Civ by Stakeholder: $F(9, 1010) = 6.41$, $p < .001$, $\eta^2 = .06$, Tukey: Prosec vs. CrimProf/CrimD-Priv/CrimD-Pub at $p < .001$, vs. Judge at $p = .001$, vs. CrimSci at $p = .002$, vs. CivLit/CivProf at $p = .004$. Private criminal defense attorneys expressed the most mean support for the use of special verdicts in civil cases ($M = 2.99$, $SD = 0.97$), but significantly more so than only prosecutors and judges. Tukey: CrimD-Priv v. Prosec at $p < .001$, vs. Judge at $p = .03$.

[xviii] Correlation between SV-Crim & SV-Civ: $r(1020) = .43$, $p < .001$.

[xix] Correlation between GV-Crim & SV-Civ: $r(1020) = -.32$, $p < .001$.

[xx] SV-Crim by Stakeholder: $F(9, 479.48) = 53.38$, $p < .001$, $\eta^2 = .23$; GV-Crim by Stakeholder: $F(9, 419.47) = 70.85$, $p < .001$, $\eta^2 = .26$.

[xxi] SV-Crim vs. GV-Crim by Stakeholder: $F(9, 1528) = 68.08$, $p < .001$, $\eta^2 = .29$.

[xxii] GV-Crim vs. SV-Crim for Prosec: $t(245) = 19.34$, $p < .001$, Cohen's $d = 1.24$; for Judge: $t(291) = 8.52$, $p < .001$, Cohen's $d = .50$.

[xxiii] GV-Crim vs. SV-Crim for LawStu: $t(46) = -6.91$, $p < .001$, Cohen's $d = -1.02$; for CrimD-Pub: $t(199) = -8.44$, $p < .001$, Cohen's $d = -60$; for CrimD-Priv: $t(248) = -6.89$, $p < .001$, Cohen's $d = -.44$; for CrimSci: $t(113) = -6.50$, $p < .001$, Cohen's $d = -.61$; for Lay: $t(136) = -2.96$, $p = .004$, Cohen's $d = -.25$; for CivProf: $t(80) = -2.95$, $p = .004$, for Cohen's $d = -.33$. Civil litigators and criminal law professors were the only two stakeholder groups who did not have significantly different mean views on general and special criminal verdicts, although civil litigators exhibited a marginal preference for special over general criminal verdicts. GV-Crim vs. SV-Crim for CivLit: $t(70) = -1.80$, $p = .08$ (marginal), Cohen's $d = -.22$.

[xxiv] TriJ on SV-Crim: $M = 2.86$, $SD = 1.19$; AppJ on SV-Crim: $M = 3.06$, $SD = 1.11$.

[xxv] Games-Howell for SV-Crim by Stakeholder: Prosec vs. all others at $p < .001$.

[xxvi] TriJ on GV-Crim: $M = 3.90$, $SD = 1.09$; AppJ on GV-Crim: $M = 3.68$, $SD = 1.08$.

[xxvii] Games-Howell for GV-Crim by Stakeholder: Prosec vs. all others at $p < .001$; Judge vs. Prosec at $p < .001$, vs. all other groups at $p < .001$.

[xxviii] Law students' and public defenders' mean views on general criminal verdicts differed significantly from those of prosecutors, judges, and jury-eligible lay citizens; the mean view of law students also differed significantly from that of criminal law professors. Games-Howell for GV-Crim by Stakeholder: LawStu vs. Prosec/Judge/Lay at $p < .01$, vs. CrimProf at $p = .002$; CrimD-Pub vs. Prosec/Judge at $p < .001$, vs. Lay at .04.

[xxix] Judges' prior litigation experience: 15% prosecution only, 9% criminal defense only, 22% civil litigation only, 5% both prosecution and criminal defense, 12% both prosecution and civil litigation, 15% both criminal defense and civil litigation, and 22% had all three types of litigation experience (prosecution, criminal defense, and civil litigation).

[xxx] SV-Crim by Judge Experience: $F(6, 407.28) = 38.18$, $p < .001$, $\eta^2 = .15$, Games-Howell: J_Pros & J_Pros-Civ vs. all other experiences at $p < .001$, except J_Pros-Civ vs. J_CrimD-Civ at $p = .004$.

[xxxi] Judges with only criminal defense experience or a combination of civil, prosecution, and criminal defense experience expressed significantly more mean support for special criminal verdicts than judges with all other types of prior litigation experience. Games-Howell for SV-Crim by Judge Experience: J_Civ vs. J_Pros/Pros-Civ at $p < .001$, vs. J_CrimD at $p = .01$, vs. J_Pros-CrimD at $p = .002$; J_CrimD & J_Pros-Civ-CrimD vs. all others at $p < .001$, except J_CrimD vs. J_Pros-CrimD at $p = .02$, J_CrimD vs. J_Civ at $p = .01$, J_Pros-Civ-CrimD vs. Civ at $p = .002$.

[xxxii] GV-Crim by Judge Experience: $F(6, 389.21) = 52.66$, $p < .001$, $\eta^2 = .18$, Games-Howell: J_Pros vs. other groups at $p < .001$, except vs. J_Pros-Civ (n/s).

[xxxiii] Games-Howell for GV-Crim by Judge Experience: J_Pros-Civ-CrimD vs. other experiences

at $p < .001$, except vs. J_Civ at $p = .02$ & vs. J_CrimD (n/s).

[xxxiv] CrimDef on GV-Crim < SV-Crim: $t(54) = -3.37$, $p = .001$, Cohen's $d = .46$.

[xxxv] GV-Crim by Role (CrimDef vs. CrimV): $F(1, 221) = 8.05$, $p = .005$, $\eta^2 = .04$.

[xxxvi] CrimJur on SV-Crim: $M = 3.51$, $SD = 1.25$; CivJur on SV-Crim: $M = 3.13$, $SD = 1.21$. SV-Crim by Juror (CrimJur vs. CivJur): $F(1, 164) = 4.07$, $p = .045$, $\eta^2 = .02$.

[xxxvii] CrimJur on GV-Crim: $M = 3.16$, $SD = 1.21$; Correlation between SV-Crim & GV-Crim for CrimJur: $r(100) = -.53$, $p < .001$.

[xxxviii] Low Severity comparisons: $t(1275) = -13.99$ vs. High Severity (Cohen's $d = -.39$), -20.99 vs. Complex (Cohen's $d = -.59$), -9.46 vs. Sex Crimes (Cohen's $d = -.27$), for -13.29 vs. Affirmative Defenses (Cohen's $d = -.37$), all at $p < .001$.

[xxxix] Complex comparisons: $t(1275) = 11.98$ vs. High Severity (Cohen's $d = .34$), 20.99 vs. Low Severity (Cohen's $d = -.59$), 16.02 vs. Sex Crimes (Cohen's $d = .45$), 8.77 vs. Affirmative Defenses (Cohen's $d = .25$), all at $p < .001$.

[xl] LawStu on Complex: $M = 4.62$, $SD = .63$. LawStu by Case Type: $F(4, 5100) = 154.75$, $p < .001$, $\eta^2 = 0.11$. LawStu on Complex comparisons: $t(39) = 4.85$ vs. High Severity (Cohen's $d = .78$), 5.59 vs. Low Severity (Cohen's $d = .90$), 7.24 vs. Sex Crimes (Cohen's $d = 1.16$), 4.49 vs. Affirmative Defenses (Cohen's $d = .72$), all at $p < .001$.

[xli] CrimD-Priv on High Severity: $M = 4.19$, $SD = 1.06$; CrimD-Pub on High Severity: $M = 4.13$, $SD = 1.03$.

[xlii] CrimD-Pub on Sex Crimes: $M = 4.05$, $SD = 1.08$.

[xliii] CrimTri on Case Type – Low Severity: $F(2, 432.04) = 88.09$, $p < .001$, $\eta^2 = .18$; High Severity: $F(2, 380.66) = 173.78$, $p < .001$, $\eta^2 = .31$; Complex: $F(2, 354.39) = 188.00$, $p < .001$, $\eta^2 = .33$; Sex Crimes: $F(2, 396.23) = 137.08$, $p < .001$, $\eta^2 = .27$; Affirmative Defenses: $F(2, 413.60) = 43.93$, $p < .001$, $\eta^2 = .11$; Games-Howell: Prosec < TriJ < CrimD for all case types at $p < .001$, except Prosec < TriJ at $p = .001$ for Affirmative Defenses. There were no significant differences between trial and appellate judges' mean views on these measures.

[xliv] Low Severity by Stakeholder: $F(9, 358.87) = 22.36$, $p < .001$, $\eta^2 = .13$, Games-Howell for Prosec vs. all other groups at $p < .001$. High Severity by Stakeholder: $F(9, 360.79) = 42.08$, $p < .001$, $\eta^2 = .24$, Games-Howell for Prosec vs. all other groups at $p < .001$. Complex by Stakeholder: $F(9, 369.50) = 49.44$, $p < .001$, $\eta^2 = .28$, Games-Howell for Prosec vs. all other groups at $p < .001$. Sex Crimes by Stakeholder: $F(9, 361.42) = 33.78$, $p < .001$, $\eta^2 = .20$, Games-Howell for Prosec vs. all other groups at $p < .001$. Affirmative Defenses by Stakeholder: $F(9, 364.40) = 18.82$, $p < .001$, $\eta^2 = .12$, Games-Howell for Prosec vs. other groups at $p < .001$, except vs. Judge at $p = .01$.

[xlv] Prosec on Affirmative Defenses comparisons: $t(182) = 5.99$ vs. Low Severity (Cohen's $d = .44$), 5.59 vs. High Severity (Cohen's $d = .42$), 3.14 vs. Complex (Cohen's $d = .23$), 5.67 vs. Sex Crimes (Cohen's $d = .42$), all at $p < .001$, except vs. Complex at $p = .002$.

[xlvi] CrimD on Affirmative Defenses comparisons: $t(366) = -6.57$ vs. High Severity (Cohen's $d = -.34$), -10.26 vs. Complex (Cohen's $d = -.54$), -3.57 vs. Sex Crimes (Cohen's $d = .19$), all at $p < .001$.

[xlvii] CrimD on Low Severity comparisons: $t(366) = -11.3$ vs. High Severity (Cohen's $d = -.59$), -13.88 vs. Complex (Cohen's $d = -.73$), -7.90 vs. Sex Crimes (Cohen's $d = -.41$), -2.87 vs. Affirmative Defenses (Cohen's $d = -.15$), all at $p < .001$, except vs. Affirmative Defenses at $p = .004$.

[xlviii] SV-Crim by Gender: $F(1, 1249.37) = 14.28$, $p < .001$, $\eta^2 = .01$; GV-Crim Gender: $F(1, 1158.80) = 19.67$, $p < .001$, $\eta^2 = .01$.

[xlix] SV-Crim by Race: $F(1, 256.53) = 9.05$, $p = .003$, $\eta^2 = .01$; GV-Crim by Race: $F(1, 1111) = 12.14$, $p < .001$, $\eta^2 = .01$. Race also exerted small but significant effects on respondents' views toward using special verdicts across different types of criminal cases. Stakeholders of color on average expressed significantly more support for using special criminal verdicts for complex, high severity, low severity, and affirmative defense cases (and marginally more support for using special verdicts for sex crimes) as compared to white respondents, but these were relatively small effects. SV-Crim for Case Type by Race – Low Severity: $F(1, 237.32) = 13.08$, $p < .001$, $\eta^2 = .01$; High Severity: $F(1, 240.01) = 5.75$, $p = .02$, $\eta^2 = .005$; Complex: $F(1, 248.57) = 9.49$, $p = .002$, $\eta^2 = .007$; Sex Crimes: $F(1, 232.93) = 4.84$, $p = .06$ (marginal), $\eta^2 = .003$; Affirmative Defenses: $F(1, 238.84)$

= 7.54, p = .007, η^2 = .006.

[l] SV-Crim by JD Status: $F(1, 576.79) = 75.35$, $p < .001$, $\eta^2 = .03$; GV-Crim by JD Status: $F(1, 496.09) = 51.58$, $p < .001$, $\eta^2 = .03$.

[li] SV-Crim by Age Category: $F(2, 684.96) = 19.48$, $p < .001$, $\eta^2 = .03$, Games-Howell: under 40 vs. 40s-50s & 60+ at $p < .001$; GV-Crim by Age Category: $F(2, 683.71) = 19.35$, $p < .001$, $\eta^2 = .03$, Games-Howell: under 40 vs. 40s-50s & 60+ at $p < .001$.

[lii] In particular, respondents from the West expressed significantly more support for special criminal verdicts, and significantly less support for general criminal verdicts, than respondents from the Midwest and South; the same was true of respondents from the Northeast as compared to respondents from the Midwest. SV-Crim by Region: $F(3, 674.11) = 16.22$, $p < .001$, $\eta^2 = .03$, Games-Howell: West vs. Midwest & South at $p < .001$, Northeast vs. Midwest at $p = .003$; GV-Crim by Region: $F(3, 674.11) = 13.58$, $p < .001$, $\eta^2 = .03$, Games-Howell: West vs. Midwest at $p < .001$ & vs. South at $p = .002$, Northeast vs. Midwest at $p < .001$.

[liii] Ideology predicting SV-Crim: $b = -.27$, $t(1239) = -9.7$, $p < .001$, $R^2 = .07$, $F(1, 1239) = 94.76$, $p < .001$; Ideology predicting GV-Crim: $b = .29$, $t(1156) = 10.32$, $p < .001$, $R^2 = .08$, $F(1, 1156) = 106.44$, $p < .001$.

[liv] GV-Crim by Prosec Race: $F(1, 152) = 10.71$, $p = .001$, $\eta^2 = .07$; GV-Crim by Judge Race: $F(1, 186) = 16.82$, $p < .001$, $\eta^2 = .08$.

[lv] SV-Crim by Judge Race: $F(1, 187) = 18.60$, $p < .001$, $\eta^2 = .09$; SV-Crim by Prosec Race: $F(1, 151) = 3.19$, $p = .08$, $\eta^2 = .02$. There was no significant difference between the opposition that white prosecutors and prosecutors of color on average expressed toward special criminal verdicts.

[lvi] Case Type by Judge Race: $F(1, 185) = 11.76$ for Low Severity, $p < .001$, $\eta^2 = .06$; 19.25 for High Severity, $p < .001$, $\eta^2 = .09$; 11.08 for Complex, $p = .001$, $\eta^2 = .06$; 13.97 for Affirmative Defenses, $p < .001$, $\eta^2 = .07$; 18.20 for Sex Crimes, $p < .001$, $\eta^2 = .09$.

[lvii] Complex by CrimD Race: $F(1, 36.20) = 7.28$, $p = .01$, $\eta^2 = .02$.

[lviii] Predicted-Prosec ($M = 1.92$, $SD = 1.17$) comparisons: vs. Predicted-CrimD: $t(1546) = -39.69$, $p < .001$, Cohen's $d = -1.01$; vs. Predicted-TriJ: $t(1546) = -21.17$, $p < .001$, Cohen's $d = -.54$; vs. Predicted-AppJ: $t(1545) = -36.15$, $p < .001$, Cohen's $d = -.92$; vs. Predicted-Jur: $t(1462) = -29.10$, $p < .001$, Cohen's $d = -.68$.

[lix] Predicted-CrimD ($M = 4.00$, $SD = 1.17$) comparisons: vs. Predicted-Prosec: $t(1546) = -39.69$, $p < .001$, Cohen's $d = -1.01$; vs. Predicted-TriJ: $t(1546) = -21.17$, $p < .001$, Cohen's $d = -.54$; vs. Predicted-AppJ: $t(1545) = -36.15$, $p < .001$, Cohen's $d = -.92$; vs. Predicted-Jur: $t(1462) = -29.10$, $p < .001$, Cohen's $d = -.68$.

[lx] Predicted-TriJ ($M = 2.60$, $SD = 1.13$) vs. Predicted-AppJ ($M = 3.28$, $SD = 1.22$): $t(1546) = -24.87$, $p < .001$, Cohen's $d = -.63$.

[lxi] TriJ on Predicted-TriJ vs. Predicted-AppJ: $t(245) = -11.03$, $p < .001$, Cohen's $d = -.71$; AppJ on Predicted-TriJ vs. Predicted-AppJ: $t(30) = -4.54$, $p < .001$, Cohen's $d = -.83$.

[lxii] Predicted-Jur: $M = 2.88$, $SD = 0.98$ (95% CI between 2.83 and 2.93).

[lxiii] Lay on SV-Crim > Lay on GV-Crim: $t(136) = 2.96$, $p = .004$, Cohen's $d = .25$.

[lxiv] Predicted-Jur by Stakeholder: $F(9, 398.98) = 6.90$, $p < .001$, $\eta^2 = .05$; Games-Howell for Prosec vs. CrimProf/CrimD-Priv/CrimD-Pub at $p < .001$, vs. Lay at $p = .003$, vs. Judge/CivProf at $p = .005$, vs. CrimSci at $p = .007$, vs. LawStu at $p = .03$, vs. CivLit at $p = .08$ (marginal); Games-Howell for Judge vs. Prosec at $p = .005$, vs. CrimD-Priv at $p = .009$. Of all the stakeholder groups, private criminal defense attorneys came closest to accurately predicting lay citizens' support for criminal special verdicts; they expected lay jurors to be significantly more supportive than prosecutors and judges did. Games-Howell: CrimD-Priv vs. Prosec at $p < .001$, vs. Judge at $p = .009$.

[lxv] Lay on SV-Crim > Lay on Predicted-Jur: $t(129) = 5.26$, $p < .001$, Cohen's $d = .46$.

[lxvi] TriJ on SV-Crim > TriJ on Predicted-TriJ: $t(246) = 5.06$, $p < .001$, Cohen's $d = .32$. Surveyed appellate judges did not exhibit this personal-group discrepancy.

[lxvii] Prosec on SV-Crim > Prosec on Predicted-Prosec: $t(248) = 3.85$, $p < .001$, Cohen's $d = .24$.

[lxviii] CrimD on SV-Crim < CrimD on Predicted-CrimD: $t(450) = -2.23$, $p = .02$, Cohen's $d = -.11$.

[lxix] CrimProf on Predicted-Prosec < CrimProf on Predicted-CrimD: $t(118) = -5.43$, $p < .001$,

Cohen's d = -.50.

[lxx]   Predicted-Prosec by CrimLaw: $F(3, 382.08) = 7.25$, $p < .001$, $\eta^2 = .02$, Games-Howell: CrimProf vs. Prosec at $p = .02$, vs. Judge at $p = .002$, vs. CrimD at $p < .001$; Predicted-CrimD by CrimLaw: $F(3, 374.15) = 13.31$, $p < .001$, $\eta^2 = .05$, Games-Howell: CrimProf vs. Prosec/Judge/CrimD at $p < .001$.

[lxxi]   Favoring in Criminal vs. Civil Cases: $X^2(6, N, 1171) = 221.57$, $p < .001$, $V = .27$.

[lxxii]   Stakeholder on Understand: $X^2(18, 1500) = 356.93$, $p < .001$, $V = .35$; Apply: $X^2(18, 1500) = 341.94$, $p < .001$, $V = .34$; Think: $X^2(18, 1417) = 287.80$, $p < .001$, $V = .32$; Discuss: $X^2(18, 1420) = 240.60$, $p < .001$, $V = .29$. LawStu ASR for Help = 5.1 on Understand, 5.0 on Apply, 3.7 on Think, 3.3 on Discuss.

[lxxiii]   Prosec ASR on Understand = 12.3 for Hinder, 7.6 for No effect, -15.2 for Help; on Apply = 13.9 for Hinder, 4.8 for No effect, -14.1 for Help; on Think = 10.0 for Hinder, 9.0 for No effect, -13.5 for Help; on Discuss = 9.4 for Hinder, 7.2 for No effect, -12.0 for Help.

[lxxiv]   Judge ASR for Help = -4.6 on Understand, -4.3 on Apply, -4.1 on Think, -4.5 on Discuss.

[lxxv]   Judge ASR for No effect = 5.1 on Understand, 4.2 on Apply, 4.6 on Think, 5.5 on Discuss.

[lxxvi]   Bias by Stakeholder: $X^2(18, 1500) = 210.07$, $p < .001$, $V = .27$.

[lxxvii]   Prosec ASR on Bias = 7.8 for Hinder, -8.9 for Help.

[lxxviii]   Convince by Litig: $X^2(8, 551) = 122.34$, $p < .001$, $V = .33$.

[lxxix]   Prosec ASR on Convince = -10.8 for Help, 5.4 for Hinder, 7.4 for No effect.

[lxxx]   Sway by Stakeholder: $X^2(18, 1268) = 145.31$, $p < .001$, $V = .24$. Prosec ASR on Sway for Decrease = -6.7; Judge ASR on Sway for Decrease = -4.1.

[lxxxi]   Prosec ASR on Sway for Increase = 8.1l.

[lxxxii]   Outcome by Stakeholder: $X^2(18, 1412) = 108.07$, $p < .001$, $V = .20$; CrimProf ASR for Convict = 4.2.

[lxxxiii]   Hung by Stakeholder: $X^2(18, 1334) = 98.72$, $p < .001$, $V = .19$; Lay ASR for Less = 6.2, Judge ASR for No effect = 4.4.

[lxxxiv]   Compromise by Stakeholder: $X^2(18, 1334) = 135.84$, $p < .001$, $V = .23$; CivProf ASR for Fewer = 7.1, Lay ASR for More = 4.4, CrimSci ASR for More = 4.7.

[lxxxv]   The survey used the terms "pro-defendant nullification" to describe nullifying acquittals ("when jurors 'nullify' the law by acquitting a defendant of a crime based on their own desired outcome, even if they think the prosecution has proven every legal element of the charged crime beyond a reasonable doubt") and "pro-prosecution nullification" to describe unproven convictions ("when jurors 'nullify' the law by convicting a defendant of a crime based on their own desired outcome, even if they do not think the prosecution has proven every legal element of the charged crime beyond a reasonable doubt"). The survey also included the broader question about "verdicts that feel morally right regardless of the law" because moral principles or perceptions of systemic unfairness could lead jurors to nullify the law even if the resulting outcome is contrary to their own desired outcomes.

[lxxxvi]   Nullifying Acquittal View by Stakeholder: $F(9, 396.63) = 179.87$, $p < .001$, $\eta^2 = .47$.

[lxxxvii]   Games-Howell for Nullifying Acquittal View by Stakeholder Group: Prosec vs. all other groups at $p < .001$.

[lxxxviii]   Nullifying Acquittal View by Judge Type (TriJ < AppJ): $F(1, 216) = 5.39$, $p = .02$, $\eta^2 = .02$.

[lxxxix]   Games-Howell for Nullifying Acquittal View by Stakeholder: CrimD-Pub vs. other groups at $p < .001$, except vs. LawStu at $p = .006$, vs. CrimD-Priv at $p = .05$ (marginal).

[xc]   Nullifying Acquittal View of CrimDef: $M = 3.74$, $SD = 1.36$; Nullifying Acquittal View by CrimParty (CrimDef > CrimV): $F(1, 216) = 8.49$, $p = .004$, $\eta^2 = .04$.

[xci]   Nullifying Acquittal View of CrimV: $M = 3.06$, $SD = 1.53$; CrimJur: $M = 3.03$, $SD = 1.51$; CivJur: $M = 2.95$, $SD = 1.51$.

[xcii]   Correlations between Nullifying Acquittal View & SV-Crim: $r(1314) = .34$, $p < .001$; Nullifying Acquittal View & GV-Crim: $r(1230) = -.37$, $p < .001$.

[xciii]   Mediation analysis indicated that respondents' views on nullifying acquittals (the independent variable) significantly predicted respondents' perceived frequency of unproven convictions (the mediator) (b = .37, t(1311) = 22.97, p < .001, R2 = .14), as well as their views on special criminal verdicts (the dependent variable) in the absence of the mediator (b = .34, t(1313)

= 37.68, p < .001, R2 = .11). Additionally, the mediator uniquely and significantly predicted the dependent variable (b = .39, t(1316) = 35.28, p < .001, R2 = .15), and the effect of the independent variable on the dependent variable shrank upon addition of the mediator into the model (Sobel z = 8.83, SE = .01, p < .001). Furthermore, support for nullifying acquittals negatively predicted perceived frequency of nullifying acquittals (b = -.26, t(1311) = 42.34, p < .001, R^2 = .07), and perceived frequency of unproven convictions negatively predicted support for general criminal verdicts (b = -.37, t(1232) = 59.88, p < .001, R^2 = .14).

[xciv] Unproven Conviction View < Nullifying Acquittal View: t(1230) = 33.88, p < .001, Cohen's d = .97.

[xcv] Unproven Conviction View < Nullifying Acquittal View for TriJ: t(195) = 8.42, p < .001, Cohen's d = 1.12; for AppJ: t(21) = 5.43, p < .001, Cohen's d = 1.04; for CrimD-Priv: t(194) = 29.08, p < .001, Cohen's d = 2.09; for CrimD-Pub: t(164) = 34.62, p < .001, Cohen's d = 270.

[xcvi] Unproven Conviction View by Stakeholder: F(9, 348.07) = 4.98, p < .001, η^2 = .05, Games-Howell: Lay vs. CrimProf/CrimD-Priv/CrimD-Pub at <.001, vs. Judge at .001, vs. Prosec at p = .002.

[xcvii] Nullifying Acquittal Frequency (M = 1.93, SD = 0.99) < Unproven Conviction Frequency (M = 2.56, SD = 1.28): t(1314) = -14.32, p < .001, Cohen's d = -.40.

[xcviii] Nullifying Acquittal Frequency by Stakeholder: F(9, 402.73) = 14.29, p < .001, η^2 = .10. Public defenders on average said nullifying acquittals occur the least frequently, and their mean rating on this measure was significantly lower than that of prosecutors, judges, and lay citizens. Games-Howell: CrimD-Pub vs. Prosec at p < .01, vs. Judge at p = .04, vs. Lay at p = .004. Prosecutors on average said nullifying acquittals occur the most frequently, and their mean rating on this measure was significantly higher than that of all the other stakeholder groups. Games-Howell: Prosec vs. all other groups at p < .001. Unproven Conviction Frequency by Stakeholder: F(9, 399.87) = 53.81, p < .001, η^2 = .22. Prosecutors on average said unproven convictions occur the least frequently, and their mean rating on this measure was significantly lower than that of all the other stakeholder groups. Games-Howell: Prosec vs. all other groups at p < .001. Public defenders on average said unproven convictions occur the most frequently, and their mean rating on this measure was significantly higher than that of all groups except private criminal defense attorneys and law students. Games-Howell: CrimD-Pub vs. Prosec/Judge/CrimProf/CivProf/Lay at p < .001, vs. CivLit at p = .02, vs. CrimSci at p = .006.

[xcix] Nullifying Acquittal Frequency vs. Unproven Conviction Frequency by Stakeholder: F(9, 1305) = 7.93, p < .001, η^2 = .05. Paired-sample t-tests for CrimD-Pub: t(178) = -16.37, p < .001, Cohen's d = -1.22; CrimD-Priv: t(205) = -14.59, p < .001, Cohen's d = -1.01; Judge: t(218) = -2.21, p = .03, Cohen's d = -.15.

[c] CrimJur on Nullifying Acquittal Frequency < Unproven Conviction Frequency: t(100) = -5.02, p < .001, Cohen's d = -.50.

[ci] Prosec on Nullifying Acquittal Frequency > Unproven Conviction Frequency: t(192) = 12.77, p < .001, Cohen's d = .92.

[cii] Unproven Conviction Frequency by Stakeholder Race: F(9, 1162) = 2.99, p = .002, η^2 = .02; Judge Race: F(1, 178) = 11.15, p = .001, η^2 = .06; CrimD-Pub Race: F(1, 165) = 4.30, p = .04, η^2 = .03; LawStu Race: F(1, 50) = 8.82, p = .005, η^2 = .15.

[ciii] Nullifying Acquittal Effect by Stakeholder: X^2(18, 1319) = 108.24, p < .001, V = .20. Prosec ASR on Nullifying Acquittal Effect = -6.6 for Decrease; Judge ASR on Nullifying Acquittal Effect = -3.2 for Decrease.

[civ] Prosec ASR on Nullifying Acquittal Effect = 5.5 for Increase.

[cv] Unproven Conviction Effect by Stakeholder: X^2(18, 1234) = 141.74, p < .001, V = .24; Lay ASR for Increase = 3.2.

[cvi] Prosec ASR on Unproven Conviction Effect for No Effect = 7.7; Judge ASR on Unproven Conviction Effect for No Effect = 4.4.

[cvii] Prosec ASR on Unproven Conviction Effect for Decrease = -7.5; Judge ASR on Unproven Conviction Effect for Decrease = -4.7.

[cviii] Nullifying Acquittal Effect vs. Unproven Conviction Effect: McNemar-Bowker X^2(3, 1232) = 89.08, p < .001, V = .35.

cix   Moral Outcomes by Stakeholder: $X^2(18, 1500) = 113.53$, $p < .001$, $V = .20$; CivProf ASR for Hinder = 3.9; LawStu ASR for Hinder = 3.0.

cx   Judge ASR on Moral Outcomes = 5.4 for No Effect; Prosec ASR on Moral Outcomes = 2.5 for No Effect.

cxi   Judge ASR on Moral Outcomes = -3.7 for Hinder.

cxii   Lay ASR on Moral Outcomes = 6.3 for Help.

cxiii   Plea by CrimTri: $X^2(6, 799) = 17.88$, $p = .007$, $V = .11$. CrimD ASR on Depends = 3.2.

cxiv   Appellate Quality by Stakeholder: $X^2(20, 1269) = 171.06$, $p < .001$, $V = .26$; Prosec ASR for No Effect = 5.9, TriJ ASR for No Effect = 5.0.

cxv   Appellate Quality by Stakeholder: Prosec ASR for Higher = -9.1, Judge ASR for Higher = -4.4.

cxvi   Appellate Quality by Stakeholder: Prosec ASR for Lower = 7.5.

cxvii   Appellate Reversal by Stakeholder: $X^2(20, 1330) = 105.04$, $p < .001$, $V = .20$; TriJ ASR for Fewer = -4.4.

cxviii   LawStu ASR on Reversal = 3.1 for Fewer.

cxix   SV-Structure by Stakeholder: $X^2(9, 1256) = 28.35$, $p < .001$, $V = .15$; Prosec ASR for Rigid = 3.3, Judge ASR for Rigid = 2.1.

cxx   Lay ASR on SV-Structure for Flexible = 3.0.

cxxi   The more politically conservative the respondents were, the more likely they were to believe that criminal jurors comply with the presumption of innocence and the less likely they were to rate criminal punishment in the American legal system as too harsh. Ideology predicting POI: $b = .31$, $t(1090) = 10.56$, $p < .001$, $R^2 = .09$; Ideology predicting Punishment: $b = -.51$, $t(1227) = -20.77$, $p < .001$, $R^2 = .26$. On average, stakeholders from the Midwest expressed significantly higher confidence in jury competence than stakeholders from the West, and significantly more faith in jurors following the presumption of innocence than stakeholders from all other geographic regions. Region on Competence: $F(3, 1170) = 3.22$, $p = .02$, $\eta^2 = .01$, Tukey: Midwest vs. West at $p = .01$; on Innocence: $F(3, 1111) = 15.17$, $p < .001$, $\eta^2 = .04$, Tukey: Midwest vs. Northeast/West at $p < .001$, vs. South at $p = .002$. Furthermore, stakeholders from the West on average regarded criminal punishment in the American legal system as harsher than did stakeholders from the Midwest and the South. Region on Punishment: $F(3, 1236) = 8.18$, $p < .001$, $\eta^2 = .02$, Tukey: West vs. Midwest at $p < .001$, vs. South at $p = .01$.

cxxii   Criminal Jury by Stakeholder: $F(9, 388.85) = 32.06$, $p < .001$, $\eta^2 = .17$; Games-Howell: LawStu vs. other groups at $p < .001$, except vs. CrimSci & Lay (n/s).

cxxiii   Games-Howell for Criminal Jury by Stakeholder: Judge vs. CivLit at $p = .04$, vs. all other groups at $p < .001$.

cxxiv   Competence by Stakeholder: $F(9, 1168) = 12.11$, $p < .001$, $\eta^2 = .17$, Tukey: Judge vs. Prosec at $p = .02$, vs. Lay/CrimSci/CrimD-Priv/CrimD-Pub/LawStu at $p < .001$, vs. CivProf at $p = .05$ (marginal), vs. CivLit at $p = .07$ (marginal).

cxxv   Tukey for Competence by Stakeholder: LawStu vs. Judge/Prosec/CrimProf at $p < .001$, vs. CivLit/CivProf at $p = .02$, vs. CrimD-Priv at $p = .09$ (marginal).

cxxvi   POI by Stakeholder: $F(9, 312.57) = 57.16$, $p < .001$, $\eta^2 = .32$, Games-Howell: Judge/Prosec vs. all other groups at $p < .001$.

cxxvii   POI by Judge Race: $F(1, 181) = 4.86$, $p = .03$, $\eta^2 = .03$.

cxxviii   POI by Judge Type (TriJ > AppJ): $F(1, 207) = 7.22$, $p = .008$, $\eta^2 = .03$.

cxxix   Games-Howell for POI by Stakeholder: CrimD-Pub vs. Judge/Prosec/CivProf at $p < .001$, vs. Lay at $p = .001$, vs. CrimProf at $p = .003$, vs. CivLit at $p = .02$.

cxxx   Punishment by Stakeholder: $F(9, 387.98) = 84.99$, $p < .001$, $\eta^2 = .39$, Games-Howell: Prosec vs. all other groups at $p < .001$.

cxxxi   Games-Howell for Punishment by Stakeholder: CrimD-Pub/CrimProf vs. Judge/Prosec/CivLit/Lay/CrimSci at $p < .001$, CrimD-Pub vs. CrimD-Priv at $p < .001$, CrimProf vs. CrimD-Priv at $p = .01$, CrimD-Pub vs. CivProf at $p = .09$ (marginal).

cxxxii   Punishment by Judge Type (TriJ < AppJ): $F(1, 201) = 5.19$, $p = .02$, $\eta^2 = .03$.

cxxxiii   Correlations for GV-Crim – Punishment: $r(1169) = -.41$, $p < .001$; Criminal Jury: $r(1177) = -.22$, $p < .001$; Competence: $r(1175) = .23$, $p < .001$; POI: $r(1117) = .45$, $p < .001$. Stakeholders

who expressed stronger support for criminal jury adjudication also tended to have more faith in criminal jurors' competence and compliance with the presumption of innocence. Correlations for Criminal Jury – Competence: $r(1177) = .38, p < .001$; POI: $r(1115) = .31, p < .001$.

[cxxxiv] Correlations for SV-Crim – Punishment: $r(1252) = .39, p < .001$; Criminal Jury: $r(1261) = -.17, p < .001$; Competence: $r(1178) = -.19, p < .001$; POI: $r(1117) = 1.41, p < .001$.