Group Processes & Intergroup Relations

G
P
I
R

Article

Group Processes & Intergroup Relations
1–18

DOI: 10.1177/1368430212441639
gpir.sagepub.com
SAGE

Is the leniency asymmetry really dead? Misinterpreting asymmetry effects in criminal jury deliberation

Norbert L. Kerr[1] and Robert J. MacCoun[2]

Abstract

Early jury simulation research, reviewed and meta-anyalysed by MacCoun and Kerr (1988), suggested a leniency asymmetry in criminal jury deliberations such that a given faction favoring acquittal will tend to have a greater chance of prevailing than would an equivalent sized faction favoring conviction. More recently, a handful of field studies of actual juries have reported either no such leniency asymmetry or one in the opposite direction (a severity asymmetry). A potential bias in the coding of these field studies' data is identified, one that would tend to underestimate any leniency asymmetry. The data from three field studies are re-analyzed after correcting this purported coding bias. The results of these re-analyses show a leniency asymmetry effect, although one that is less pronounced than observed in mock jury studies. It is argued that this difference in degree (not existence) of leniency asymmetry can plausibly be attributed to greater imbalance in evidence strength in the typical actual trial relative to the typical stimulus case in simulation experiments. It is also noted that failure to observe such a leniency asymmetry effect in actual juries would raise important questions about their adherence to the reasonable doubt standard of proof.

Keywords

deliberation, jury, jury simulation, leniency asymmetry

Paper received 26 July 2011; revised version accepted 09 February 2012.

During the late 1970s and early 1980s, a number of investigators reported finding a "leniency bias" in mock jury deliberations involving dichotomous (Guilty vs. Not Guilty) verdict options, such that juries collectively show a stronger propensity to acquit the defendant than one would predict based on the distribution of jurors' predeliberation opinions (e.g., Davis, Stasser, Spitzer, & Holt, 1976; Kerr et al., 1976; Stasser, Kerr, & Bray, 1982; Kerr & MacCoun, 2007). In a 1988 paper, we showed that this effect was quite robust in a meta-analysis of 11 mock criminal jury studies. Then, in a new experiment, we demonstrated that the bias was not simply attributable to lenient attitudes among the typical college-student mock jurors, but was

[1]Michigan State University, USA and University of Kent, UK
[2]University of California, USA

Corresponding author:
Norbert L. Kerr, Michigan State University, Department of Psychology, East Lansing, MI 48824, USA.
Email: kerr@msu.edu

also observable in a non-student community adult sample. In a final experiment, we showed that the bias was moderated by the standard-of-proof instructions the mock jurors received—viz. the bias was eliminated when mock jurors received preponderance of evidence instructions rather than reasonable doubt instructions routinely used in criminal trials (also see Kerr et al., 1976; MacCoun, 1984; Stawiski, Dykema-Engblade, & Tindale, 2012).

The label "leniency bias" accurately characterizes the empirical phenomenon discussed in our 1988 paper, but in hindsight, we can see that the term also lends itself to misinterpretation. First, to the degree that it stems from criminal jurors following the court's instructions on the reasonable-doubt standard, perhaps it should not have been termed a "bias." Second, it does not refer to a general tendency for jurors or juries to be lenient. Rather, it refers to a specific asymmetry such that a given faction favoring acquittal will tend to have a greater chance of prevailing than would an equivalent sized faction favoring conviction. For both reasons, we will hereafter refer to this departure from symmetry in juries' social decision scheme (Davis, 1973) as the leniency asymmetry effect.

From our perspective, this effect is less significant as an empirical regularity than as a source of insights into jury dynamics. A simple and convenient assumption, made by the Supreme Court in Williams v. Florida (1970), is that a jury faction's influence is strictly proportional to its size; thus, both factions should have the same chance of prevailing in an evenly split 6:6 group, a 8:4 split should have a 66% likelihood of producing a conviction, a 4:8 split should have a 66% likelihood of producing an acquittal, and so forth. This proportionality assumption has been decisively rejected by empirical research on mock and actual juries; instead, majorities have influence disproportionately greater than their relative size would predict. The leniency asymmetry qualifies this majority amplification effect, showing that the disproportional power of a majority in a criminal trial jury is stronger when it favors acquittal rather than conviction, and that equally split juries are relatively more likely to acquit than to convict.

Initial rumors of the leniency asymmetry's demise

In their review of the jury decision making literature, Devine, Clayton, Dunford, Seying, and Pryce (2001) first raised questions about the robustness and generality of the leniency asymmetry effect. After reproducing the meta-analysis and conclusions of MacCoun and Kerr (1988) based on mock criminal juries, they noted that a pair of studies that permitted an examination of leniency asymmetry biases in actual criminal juries did not appear to reveal such an effect (Kalven & Zeisel, 1966; Sandys & Dillehay, 1995; also see Zeisel & Diamond, 1978, for a recasting of the Kalven & Zeisel, 1966, data). Specifically, Devine et al. asserted that "…Kalven and Zeisel (1966) found a 50% conviction rate in their 10 evenly divided juries, whereas Sandys and Dillehay found a 71% conviction rate in the 24 juries that were evenly split juries on the first ballot" (p. 693).

Both studies used post-trial interviews with jurors to estimate the first ballot vote. Unfortunately, the specific methods used by Kalven and Zeisel are sketchy (it is referred to as "a pilot study" and no details are provided), but Sandys and Dillehay describe their own procedures in more detail. They strove to interview three of the 12 jurors in each of 50 focal criminal trials, and required that two of the interviewed jurors had to agree on the classification of the jury's first ballot vote (using Kalven and Zeizel's five categories—unanimous for not guilty; minority of guilty votes; six guilty votes; majority of guilty votes; and unanimous for guilty). Even for these broad categories, no such agreement was obtained in seven (14%) of the focal trials. The collapsing of all juries with and without majorities of guilty votes precludes specific row by row comparisons (e.g., 8:4 vs. 4:8 G:NG ratios), but the juries with six Guilty votes could be informative. There were only three such juries in the 43 classifiable focal juries and all returned Guilty verdicts (100% conviction rate). The statistic noted by Devine et al. (viz. 71%) was based on a larger sample of trials including 190 non-focal trials (juries on which

Table 1. Summary of reported leniency asymmetry tests in actual juries, according to source authors' coding schemes for first-ballot votes.

Data Source	Number of juries with half the jurors voting for guilty at first ballot	Reported relative frequency (Percentage) acquitting [excluding hung juries] using source authors' coding	95% Confidence Interval**
Kalven & Zeizel (1966)	10	5/10 (50%)	23.4%–76.6%
Sandys & Dillehay (1995; focal juries)	3	0/3 (0%)	0%–52.7%
Sandys & Dillehay (1995; focal and nonfocal juries)	24	5/22 (22.7%)	9.1%–42.1%
Devine et al. (2004; 6-person juries)	5	1/5 (20%)	4.3%–64.1%
Devine et al. (2004; 12-person juries)	4	2/4 (50%)	14.7%–85.3%
Devine et al. (2007) original coding assumptions	7	1/6 (16.7%)	3.7%–57.9%
Hannaford-Agor et al. (2003)	44	21/44 (47.7%)	33.7%–62.1%

**Computed using an exact Bayesian calculator (http://www.causascientia.org/math_stat/ProportionCI.html)

respondents reported serving during the same period during which the focal trials occurred), where the recollection of a single juror was deemed sufficient to classify the first-ballot vote. Of the 233 total focal and non-focal juries, 24 fell in the "six guilty votes on the first ballot" category, of which 17 convicted, five acquitted, and two hung. We summarize all of the relevant data in the first three few rows of Table 1, which makes clear that these two studies suggest either no asymmetry (Kalven & Zeisel) or a severity asymmetry (Sandys & Dillehay). Neither is consistent with the strong leniency asymmetry summarized in MacCoun and Kerr (1988), which reported a weighted average p (Acquit | start with even split and reach a verdict) = .808; the comparable figures are .50 for the Kalven and Zeisel data, and either .00 or .23 for the Sandys and Dillehay paper (depending on which juries are included). Only for the larger, Sandys and Dillehay data set (including both focal and non-focal juries) can a leniency asymmetry effect (i.e., p (Acquit | start with even split and reach a verdict) >.50) be statistically rejected. These results led Devine et al. (20011) to draw the conclusion that "Overall, although there is clearly not yet enough data to draw a definitive conclusion, the strong leniency bias observed in laboratory studies may be weaker or less reliable in actual juries" (p. 693).

In a pair of subsequent studies (Devine et al., 2004; Devine, Buddenbaum, Houp, Studebaker, & Stolle, 2007), Devine and his colleagues responded to their earlier call for more data on this question. Devine et al. (2004) surveyed criminal trial jurors immediately after their trials. There were 97 juries for which data were obtained for at least two jurors; however, for only 79 of these juries was there sufficient agreement among respondents to estimate a first-ballot verdict split (i.e., for approximately 32% of the juries, jurors could not agree on the first ballot vote, even when they were surveyed immediately after that vote).[1] The relevant data for these 79 juries are reproduced in Table 2. For our purposes, the key data are what Devine et al. called "close" juries

Table 2. Data relevant to a leniency asymmetry effect from Devine et al. (2004).

Vote 1 preference distribution	Final verdict		
	Not guilty	Hung	Guilty
6-persoii juries (*N*= 22)			
0 G, 6 NG	6	—	1
1 G, 5 NG	2	—	—
2 G, 4 NG	1	—	—
3 G, 3 NG	1	—	4
4 G, 2 NG	—	1	—
5 G, 1 NG	—	—	—
6 G, 0 NG	—	—	6
12-person juries (*N* = 57)			
0 G, 12 NG	13	—	—
1 G, 11 NG	—	—	—
2 G, 10 NG	4	1	—
3 G, 9 NG	2	—	—
4 G, 8 NG	3	—	—
5 G, 7 NG	*1*	—	*3*
6 G, 6 NG	*2*	—	*2*
7 G, 5 NG	—	—	*3*
8 G, 4 NG	—	1	4
9 G, 3 NG	—	—	3
10 G, 2 NG	—	—	4
11 G, 1 NG	—	1	5
12 G, 0 NG	—	—	5

*Note: N= 79 juries***.** Table values represent frequency counts. Italicized distributions are those without a strong (i.e., 2/3) majority. G = number of jurors voting for guilty; NG = number of jurors voting for not guilty. The table has been Reprinted with permission.

(i.e., those with less than a 2/3 majority). They summarize the data for such juries as follows:

> Surprisingly, instead of acquitting, 75% (i.e., 12 of 16) of the close juries ended up convicting the defendant… . Thus, there was no indication of a leniency bias favoring defendants in the absence of a strong two-thirds majority on the first deliberation vote. (p. 2081)

They subsequently conclude that their data document an opposite asymmetry effect, a "severity bias." For purposes of comparison, we summarize the data for Devine et al.'s (2004) "equal-split" juries in Table 1. As with the earlier studies, there is no indication of a leniency asymmetry effect, but such an effect cannot be statistically ruled out (i.e., the 95% Confidence Interval, presented in the final column, a range of values exceeding .50).

Devine et al. (2007) used a roughly similar methodology[2] to survey members of 103 criminal juries (14 six-person and 89 12-person juries) for which sufficient agreement on first-ballot votes on the first/major charge could be obtained. They presented their results in terms of the percentage of first ballot guilty votes (i.e., combining the six- and 12-person juries); their tabled results are reproduced in Table 3. Again, they focus on "close" juries (i.e., those with .33 < (proportion favoring conviction) < .67), and note that most of these (10 out of 18 or 56%) convicted. They took this as strong evidence against a leniency asymmetry effect. The most potentially informative data are again in the juries where 50% favored conviction. As shown in the next to last row of Table 1, the acquittal rate was quite low (16.7%), which is consistent with a severity asymmetry effect (although again, the statistical power was inadequate to statistically reject a leniency asymmetry effect).

Finally, in an ambitious study of hung juries, Hannaford-Agor, Hans, Mott, & Munsterman (2002; also see Hans, Hannaford-Agor, Mott, & Munsterman, 2003) had jurors from 382 trials in four different jurisdictions report on the first ballot splits in their juries. As in Kalven and Zeisel (1966), they collapsed certain ranges of initial splits (see Fig. 1) on the trial's first count/charge. Again, such collapsing precluded detailed row by row comparisons (e.g., 8:4 vs. 4:8 G:NG ratios), but again, the juries with six Guilty votes could be informative for detecting asymmetry. They report 44 juries with such even splits, of which 23 (52.3%) ultimately convicted; yet again, the nonsignificant trend (see Table 1) was toward a severity asymmetry effect.

Potential biases in coding of predeliberation splits

There are several potential explanations for the apparent divergence of laboratory and field evidence on the leniency asymmetry (e.g., see Note 4, *Infra*).

Table 3. Data relevant to a leniency asymmetry effect from Devine et al. (2007).

Proportion of Jurors Favoring Conviction, Vote 1	Final Jury Verdict			
	Conviction	Hung	Acquittal	Conviction Rate
0.00	0	0	12	0%
0.08	0	0	2	0%
0.17	1*	0	6	14%
0.25	0	2	2	0%
0.33	1	0	1	50%
0.42	0	1	0	0%
0.50	5	1	1	71%
0.58	5	2	3	50%
0.67	4	4	0	50%
0.75	5	0	0	100%
0.83	14	0	0	100%
0.92	4	1	0	80%
1.00	26	0	0	100%
Totals	65	11	27	103

Note: Numerical values in the middle columns represent the number of juries arriving at the respective verdict indicated in the column heading given the proportion of jurors favouring conviction on the first vote indicated in the left-hand column. *N*=103 criminal juries. * Indicates a jury convicted on a lesser charege. The table has been Reprinted with permission.

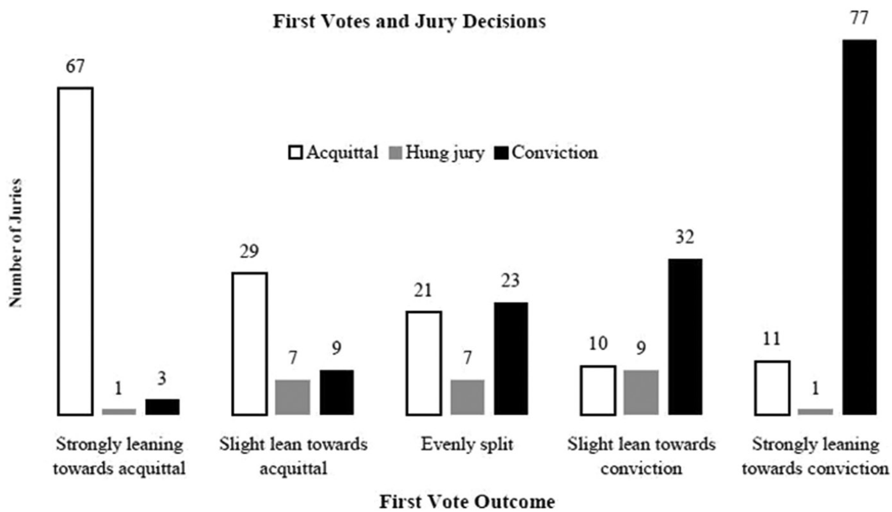

Figure 1. Plot of initial-ballot to final-verdict distributions reproduced from Hannaford-Agor et al. (2002).
Note: Reprinted with permission

We suspect that one—perhaps the primary—explanation may lie in an apparently innocuous but crucial assumption made in the coding of the field studies' data. Devine et al. (2004) state (and justify) this assumption as follows: " … we collapsed the *not guilty* and *uncertain/abstain* categories based on the notion that individuals in the latter category were indicating through their behavior that they possessed reasonable doubt." Thus, Devine et al. (2004, 2007) functionally equated either a vote of "undecided" at the first ballot or an abstention from that first ballot as a preference for a Not Guilty vote at the start of deliberation. Inspection of Sandys and Dillehay's (1995) and Kalven and Zeisel's (1966) methods indicates that they made the same coding assumption as Devine et al. (2004, 2007).[3]

What is the net effect of this assumption? Let's consider an example. Suppose that prior to deliberations, there were five jurors who favored conviction, two that were undecided or would abstain at the first ballot, and five who favored acquittal; we might summarize this as $(N_G, N_{U/A}, N_{NG})$ = (5,2,5). In the typical laboratory study, participants are not given the choice of undecided or abstain (although a very few do decline to respond) but must indicate a preference for either a Guilty or a Not Guilty verdict; it is the association between the distribution of such forced-choice predeliberation preferences and (mock) jury verdicts that constituted the core evidence for the leniency asymmetry effect. The key question, then, is how would the two undecided/abstaining jurors respond if they were required to make a choice? Devine et al. assume that they both would vote for Not Guilty, resulting in a predeliberation split of $[N_G, N_{NG}]$ = [5,7] = 58.3%G. But is there a good empirical or logical reason to suppose that jurors who say they are undecided really are not? We know of none, and thus, would suggest that one should either drop the undecideds from such an analysis {i.e., (5,2,5) → [5,5]} or assume that the undecideds are equally likely to vote for G or NG {i.e., (5,2,5) → [6,6] }; in either case, (5,2,5) → 50% guilty. Likewise, there are several reasons why a juror might choose to abstain in an initial jury ballot—e.g., s/he might want to avoid joining any faction before s/he has heard more deliberation; s/he may want to avoid majority pressuring as long as possible; s/he may object in principle to the method of balloting (e.g., failure to use a secret ballot). We know of no evidence, though, suggesting that such motives are more likely to characterize those who are leaning toward an acquittal than those who are leaning toward a conviction. So again, we suggest that a more defensible approach would be either to exclude or evenly-split such abstentions. Thus, a hypothetical jury that under the most defensible assumptions should be coded as equally-split between G and NG factions (50% Guilty), would under Devine et al.'s assumption be classified as pro-acquittal (58.3% Guilty). This is the general effect of Devine et al.'s coding assumption—to misclassify juries as more pro-acquittal than they probably are.[4]

Let's consider a second, telling example. Suppose the initial distribution was $(N_G, N_{U/A}, N_{NG})$ = (6,4,2). Under Devine et al.'s assumption, this would be classified as an "equal-split" jury, $[N_G, N_{NG}]$ = [6,6] = 50.0%G, and of particular interest for detecting an asymmetry in jury decision making. But if one either dropped the undecideds {$[N_G, N_{NG}]$ = [6,2] = 75.0%G} or split them equally {$[N_G, N_{NG}]$ = [8,4] = 67.7%G}, one would classify this as a fairly strongly pro-conviction jury. In principle then, some (or many?) of Devine et al.'s (2004, 2007) "equal-split" juries may in fact have begun deliberation with majorities favoring conviction, and hence, have been inappropriate for estimating the parameter that is crucial for the leniency asymmetry effect (viz. *p* (Acquit|start with even split and reach a verdict).

Fortunately, there is potentially a way to empirically resolve these dueling coding assumptions—one might recode these data under different coding assumptions to see if it would alter the degree of leniency vs. severity asymmetry. We first contacted Prof. D. Devine and he sent us the data files for both Devine et al. (2004) and Devine et al. (2007). Unfortunately, first-ballot undecided or abstain votes were not coded as such for Devine et al. (2004) and the raw data were no longer available (D. Devine, personal communications, 15 May 2009 and 8 July 2009). However, the coding for the Devine et al. (2007) data did retain this information. Additionally, Prof. R. Dillehay supplied the data for the 50 focal juries examined in Sandys and Dillehay (1995). Finally, Paula Hannaford-Agor provided a data deck for the large sample of trials examined in Hannaford-Agor et al. (2002). Thus, it was possible for us to reanalyse the data for these three studies under alternative coding assumptions.

Reanalyses of Devine et al. (2007)[5] Sandys and Dillehay (1995),[6] and Hannaford-Agor et al. (2002)[7]

In Table 4 are three alternative codings of the data from these three field studies. In the first column is the percentage of the jury voting for conviction on the first ballot. The remaining columns present the distribution of jury verdicts for various initial ballot splits under three coding assumptions:

1. All undecided or abstentions on the first ballot are assumed to favor Not Guilty (the assumption made in prior field studies).
2. All undecided and abstentions are split equally between the factions favoring Guilty and Not Guilty. Occasionally, this results in fractional first ballot votes. For example, one 12-person jury had a vote of 11 for Not Guilty and one Undecided at their first ballot. This jury was classified as having 0% for Guilty under Devine et al.'s coding assumption (#1. Un&Ab = NG). Under the present coding assumption, it was assumed to have 11.5 votes for Not Guilty and .5 for Guilty, and hence, .5/12 = 4.2% for Guilty. Such cases sometimes resulted in new rows in the table that did not occur under the other assumptions examined (for which there were no fractional votes). Also, since some of the juries in Hannaford-Agor et al.'s study were eight-person juries, some rows (e.g., for 1G:7NG splits, for which %G = 12.5%) only occur for this study.
3. All juries with any undecided or abstaining voters at the first ballot were dropped from the analysis. Naturally, this reduces the size of the data sets (across studies, between 20% to 40% of the juries contained some undecideds or abstains). The patterns of the data under this third coding assumption are similar to those under the preceding assumption (#2); the data are provided here primarily for the sake of completeness.

Focusing on the first two coding alternatives, several patterns are evident from visual inspection. For both codings, initial majorities nearly always prevail (in 191 of 205 [93.2%] juries under coding assumption #1; in 205 of 218 [94.0%] under coding assumption #2). Consistent with a leniency asymmetry effect, the exceptions are more likely to be pro-conviction majorities losing, and this is more consistently the case under coding assumption #2 (in 12 of 13 instances), which we have argued provides less biased estimates of initial juror preference than under coding assumption #1 (in nine of 14 instances), which we have suggested biases estimates of initial juror preference. In the "50% guilty" row, one sees that the severity asymmetry more pronounced under coding assumption #1 (11 of 12 verdicts being convictions) than under the more defensible assumption #2 (for which three of four verdicts were convictions). Considering the entire matrices, the severity asymmetry reported when coding assumption #1 is used is attenuated or eliminated when coding #2 (or #3) is employed.

However, we require a better way of probing for asymmetry, one that takes into account all the available data rather than focusing on isolated rows in the table (e.g., the single, equal-split row, for which the available sample sizes are very small and hence, the statistical power is very low; see Table 1). Estimating and plotting the relationship between first ballot vote and final verdict would accomplish this. Such a plot is essentially a graphical representation of the operative *social decision scheme* matrix for juries [excluding hung juries] (cf. Davis, 1973; Kerr, MacCoun, & Kramer, 1996; Stasser, 1999). Zeisel and Diamond (1978) attempted this for Kalven and Zeisel's (1966) data; their result is reproduced in Figure 2a. The x-axis is the % of guilty votes on the first ballot. The solid squares are the observed % of convictions (scaled on the right-hand y-axis) for Kalven and Zeizel's five first-split categories (viz.

Table 4. Old and new codings of Devine et al. (2007), Sandys & Dillehay (1995), and Hannaford-Agor et al. data.

	Field Study																	
	Devine et al. (2007)						Sandys & Dillehey(1995)						Hannaford-Agor et al. (2002)					
	Coding Assumptions																	
	1.Un&Ab=NG		2. Un&Ab=half G/half NG		3.drop juries with Un&Ab		1.Un&Ab=NG		2. Un&Ab=half G/half NG		3.drop juries with Un&Ab		1.Un&Ab=NG		2. Un&Ab=half G/half NG		3.drop juries with Un&Ab	
	Final Jury Verdict																	
1st ballot %G	G	NG	G	NG	G	NG	G	NG	G	NG	G	NG	G	NG	G	NG	G	NG
0.0%	0	12	0	11	0	11	0	6	0	5	0	5	0	10	0	9	0	9
4.2%			0	1														
8.3%	0	2	0	2	0	2	0	0	0	1	0	0			0	6	0	4
12.5%													0	2	0	1	0	1
16.7%	0	7	0	5	0	5	0	0	0	0	0	0	0	1	0	2	0	0
25.0%	0	2	0	2	0	2	0	1	0	1			0	3	0	3	0	2
33.3%	1	1	0	1	0	1	0	1	0	1	0	1	2	3	1	3	0	2
41.7%	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	1		
50.0%	5	1	1	1	1	1	2	0	1	0	1	0	3	0	1	0	1	0
54.2%			1	0														
58.3%	5	3	4	5	4	3	0	0	1	1	0	0	3	0	5	1	1	0
62.5%			2	0									3	0	3	0	2	0
66.7%	4	0	4	0	3	0	2	1	0	0	0	0	7	1	5	1	3	2
75.0%	5	0	4	0	2	0	2	0	2	0	0	0	9	0	10	0	7	0
79.2%			2	0														
83.3%	14	0	8	0	8	0	6	2	6	2	3	2	4	2	4	2	1	2
87.5%			2	0									4	0	3	0	3	0
91.7%	4	0	8	0	2	0	7	0	10	0	2	0	3	0	5	0	1	0
95.8%			2	0														
100.0%	26	0	26	0	26	0	14	0	13	0	13	0	17	0	20	0	17	0
Total	64	28	64	28	46	25	33	11	33	11	19	8	57	23	57	29	36	22
	Estimates of bBOP model parameters																	
	b= 0.42 c= 9.81		b= 0.55 c= 15.5		b= 0.54 c= 14.9		b= 0.46 c= 6.54		b= 0.52 c= 7.14		b= 0.57 c= 5.78		b= 0.38 c= 18.6		b= 0.47 c= 14.5		b= 0.49 c= 8.39	

Notes: Un = Undecided on 1st ballot. Ab = Abstained on 1st ballot. G = Guilty. NG = Not guilty.

unanimous for not guilty, minority of guilty votes, six guilty votes, majority of guilty votes, and unanimous for guilty; for the 2nd and 4th categories we take the midpoint of the range covered as the abcissa value). The solid dots are the observed % of acquittals (scaled on the left-hand y-axis). The two curves shown are Zeisel and Diamond's "freehand interpolations" of these data points—the upper curve estimating the probability of the jury acquitting (scaled against the left-hand y-axis) and the lower estimating the probability of the jury convicting (scaled against the right-hand y-axis). The distance between these two curves was an estimate of the proportion of hung juries expected for any initial split; the closeness of the two curves reflect the generally low hung jury rate in these data. By visual inspection, these curves seem fairly symmetric (i.e., factions for conviction are assumed to be as likely to prevail as like-sized factions for acquittal). This reflects both the data, which show no marked asymmetry in either direction, and the historical fact that conceptual and empirical speculation about any such asymmetry arose more than a decade after *The American Jury* was published (Davis et al., 1976; MacCoun & Kerr, 1988)—at the time, there was little reason for Diamond and Zeisel (1978) to be looking for even subtle indications of asymmetry.

Such freehand curve fitting was adequate for Zeisel and Diamond's immediate objectives, but our present objectives require more precise estimation methods. Ideally, such a method should be responsive to all data points, should utilize a quantitative goodness-of-fit criterion, should offer a simple way to detect asymmetry in the function, and should have a defensible theoretical rationale. Fortunately, MacCoun (in press) has recently provided just such a function and estimation method. MacCoun suggests that the utilization of subjective judgmental thresholds by both individuals and groups routinely produces a sigmoid (S-shaped) function relating the latent propensity to give a response (e.g., the strength of evidence against a defendant) and the likelihood or strength of response (e.g., the probability of conviction), much like the S-shaped function that describes Kalven and Zeisel's data in Figure 2a. He derives a family of logistic-threshold models of social influence called the *BOP* (Balance of Pressures) models. He then applies the BOP models to describe classic findings in a variety of social influence settings including unidirectional influence [e.g., persuasion and conformity], imitation, and bidirectional influence [e.g., jury deliberation] and shows that they provide a superior qualitative and quantitative fit than competing models. Of particular interest for us, he takes Kerr and MacCoun's (1985) 12-person mock jury data and shows that the bBOP model (bidirectional BOP model) provides an excellent fit to their observed relationship between initial splits and final mock jury verdicts.

Using our current terminology, the bBOP model may be expressed as

Probability that jury convicts = $\{ 1 + \exp$ (-c * [(proportion of jurors initially favoring conviction)-b]) $\}^{-1}$

where c and b are two free model parameters. Importantly, these parameters have substantive meaning. As c, the *norm clarity* parameter, approaches its lower limit ($c > 0$), the S-shaped function becomes "flatter" (less inflected). As c increases, the S-function becomes more sharply inflected; as $c \rightarrow \infty$, the function approaches a step-function. Thus, the magnitude of c suggests how big an impact is felt by approaching and passing some judgmental threshold. The b (*burden of proof*) parameter locates that threshold on the dimension of judgment (in our case, the % of jurors voting for conviction). It is the point where a small change along that dimension has the biggest effect on the jury outcome (i.e., it is the inflection/breaking point of the S-shaped function, where the slope of the function is the steepest). In the current case, this means that if $b = .50$, the midpoint of the initial-split dimension, the response function is symmetric and advocates of conviction are no more or no less able to prevail than advocates of acquittal. If $b > .50$, this means that the threshold, the perfect balance between a pro-conviction faction and a pro-acquittal faction, occurs when the pro-conviction fac-

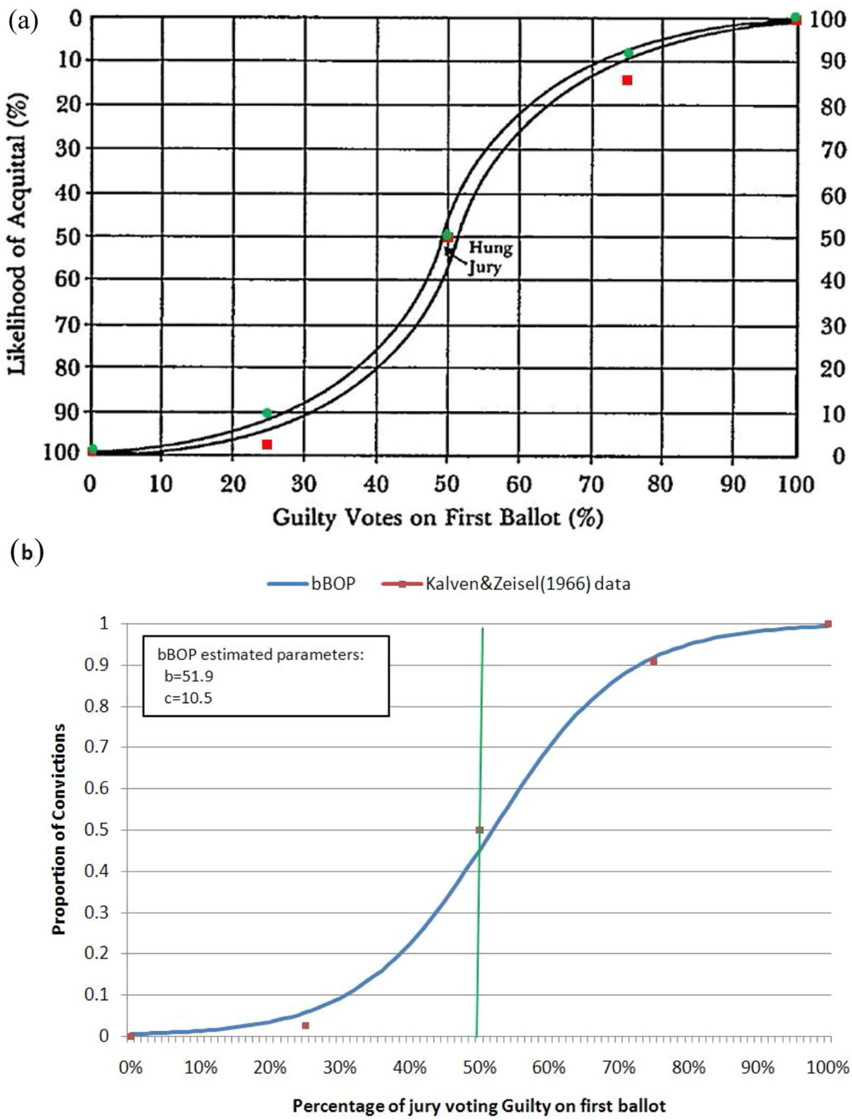

Figure 2. Estimation of the decision scheme implied by Kalven and Zeisel's (1966) data.
(a) Zeisel and Diamond's (1978) freehand estimation. The solid squares are the observed % of convictions (scaled on the right-hand y-axis). The solid dots are the observed % of acquittals (scaled on the left-hand y-axis).
(b) Best fitting bBOP model of conviction rate data (dropping hung juries).

tion = b, and hence, it requires more than half of the jury to favor a conviction before the conviction rates shift quickly from pro-acquittal to pro-conviction. The response function is asymmetric, with a leniency asymmetry. Conversely, when b < .50, the reverse is true, and the function exhibits a severity asymmetry.

Let's illustrate this concretely. In Figure 2b we show the result of fitting the bBOP model to Kalven and Zeisel's (1966) conviction data (the squares). Because the leniency asymmetry hypothesis applies only to juries that reach a verdict, we focused simply on the rate of convictions, after dropping hung juries. The bBOP model was fit by minimizing the weighted mean squared error (using the relative frequency of each data point as a weight) using Excel's Solver function with the quasi-Newton algorithm and quadratic extrapolation. The bBOB function tells a similar story to that told by Zeisel and Diamond's freehand fit—the function is nearly symmetric (b = 51.9%).[8] The statistic that MacCoun & Kerr (1988) focused upon as the key index of the leniency asymmetry was P(jury votes guilty | jury is evenly split originally and reaches a verdict). In the figure, that value is the point at which the curve intercepts the vertical line at 50% of jurors voting for G on the first ballot (the vertical line in the figure). Here, that estimate is .45 (indicative of a weak leniency asymmetry), but quite different than the mean estimate from MacCoun & Kerr's (1988) meta-analysis of mock jury studies (viz. .192, indicating a very strong leniency asymmetry). Like all the other field studies, Kalven and Zeisel lumped first-ballot undecideds together with acquittals which we have argued biases the estimates of first ballot splits. This bias—plus a possible bias in assuming that first ballot votes are identical to pre-deliberation verdict preferences (see Note 4, *Supra*)—suggests that this function may very well misestimate the presence and magnitude of any asymmetry in Kalven and Zeisel's juries.

We can eliminate at least the first of these two biases (viz. the purported coding bias) in the three field studies that have been recoded in Table 4. In Figure 3 we present the result of fitting the bBOP model to Devine et al.'s (2007) data under the original coding assumption (#1, counting all undecideds and abstentions as Not Guilty votes) and our recommended alternative coding assumption (#2, splitting all undecided and abstain votes equally between Guilty and Not Guilty). Figure 3a confirms Devine et al.'s summary of their data (under assumption #1)—the best fitting function is asymmetric in the direction of a severity asymmetry (b = 42.2%, with an estimated P(convict | even initial split) = .683. The symmetric version of the bBOB function, with b = 50%, is also plotted (the dashed curve) for purposes of comparison. A severity asymmetry occurs whenever the best fitting function is above/to the left of this symmetric curve; a leniency asymmetry occurs whenever the best fitting function is below/to the right of the symmetric curve. Figure 3b shows that the asymmetry is reversed under a more defensible coding assumption—now the data are best fit by a function with a leniency asymmetry (b = 55.0%, with an estimated *p* (convict | even initial split) = .316). Much the same leniency asymmetry occurs under the final, most conservative coding assumption (#3, drop all juries with undecided or abstentions at ballot one), viz. b = 54.1% and the estimated *p* (convict | even initial split) = .353. The pattern is similar (if slightly less dramatic) when the same bBOP fitting is applied to the other two field studies (Hannaford-Agor et al., 2002; Sandys & Dillehay, 1995; see the bottom of Table 4 for all estimated model parameters). For each, if all undecideds are counted as NG votes (coding assumption #1), there is a clear severity asymmetry (estimated b's were .46 and .38, respectively). If undecideds are split equally between G and NG factions (coding assumption #2), then the best fitting bBOP curves are nearly symmetric (b = .47 for Hannaford-Agor et al.; b = .52 for Sandys & Dillehay). And if all undecideds are dropped (coding assumption #3), again the best fitting bBOP curve is nearly symmetric (b = .49 for Hannaford-Agor et al.) or shows a leniency asymmetry (b = .57 for Sandys & Dillehay).

In summary, the severity asymmetry reported in several studies of actual juries can plausibly be attributed, at least in part, to a biased coding assumption. When that bias is removed, the data reveal either little asymmetry (for one study) or a leniency asymmetry (in two other studies). One remaining question is why the latter leniency asymmetry is weaker (e.g., *p* (convict | even initial split) ≈ .33 for Devine et al.) than the effect suggested by MacCoun and Kerr's (1988) meta-anaysis (i.e., *p*(convict | even initial split) ≈ .20).

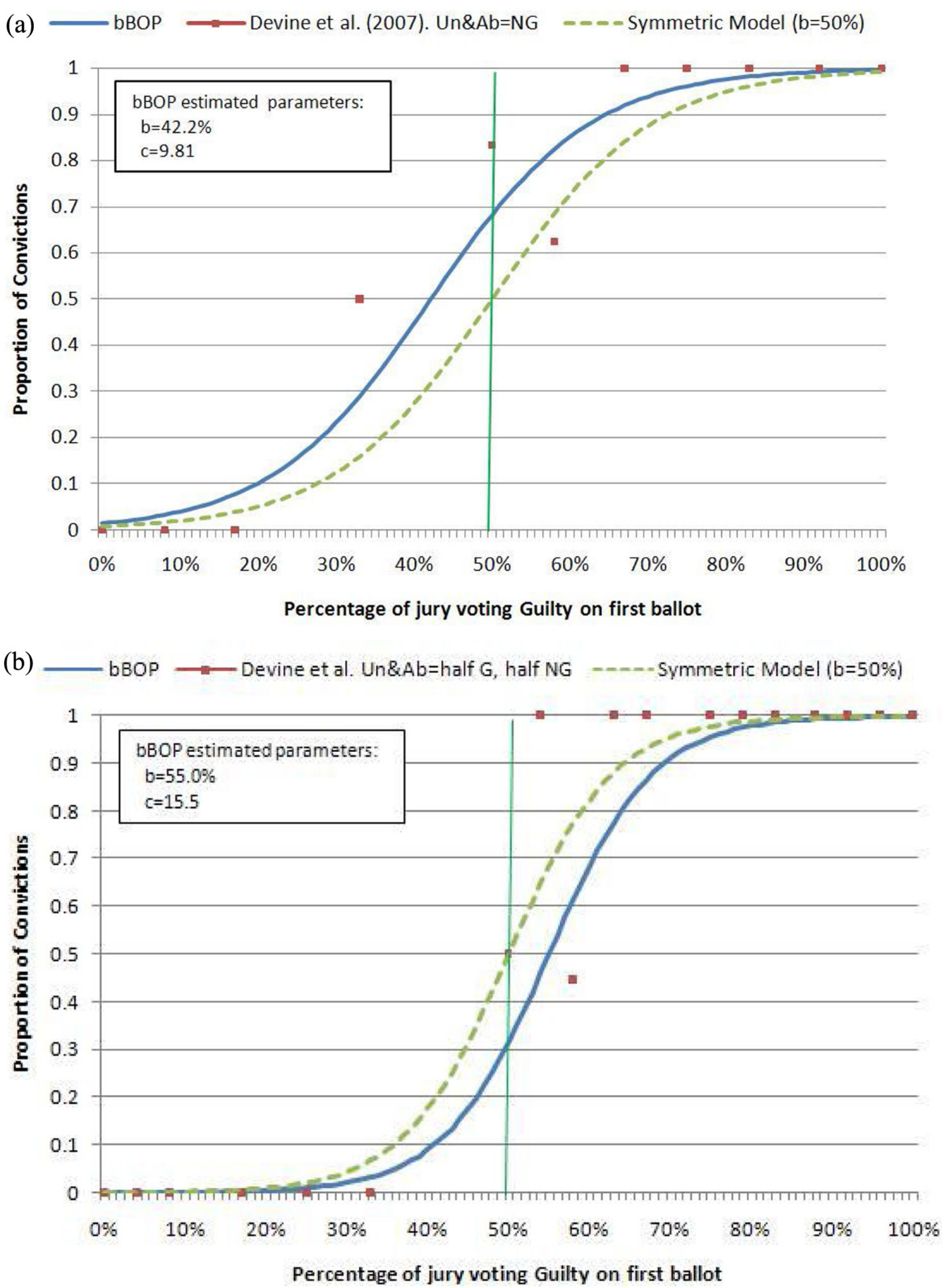

Figure 3. Best fitting bBOP models of Devine et al.'s conviction rate data.
(a) Under coding assumption (#1) that undecideds and abstentions are all votes for Not Guilty.
(b) Under coding assumption (#2) that undecideds and abstentions are half Guilty, half Not Guilty.

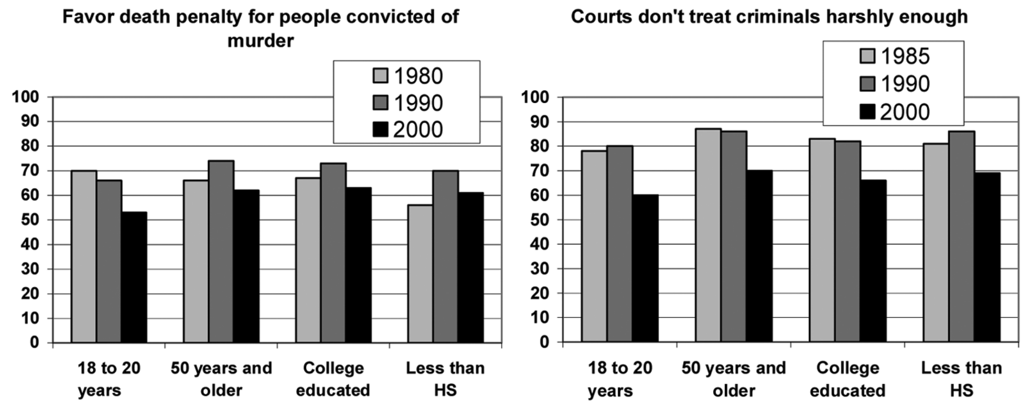

Figure 4. Trends in age and education differences in punitiveness for crime.
Source: Sourcebook of Criminal Justice Statistics Online (2009a, 2009b).

Is there any reason why mock juries should be more lenient than real juries?

For many readers, we suspect that any suggestion of a discrepancy between “real” vs. “mock” juries will evoke longstanding arguments about the alleged invalidity of the mock jury method — arguments we review and critique elsewhere (Kerr & Bray, 2005; MacCoun, 2005). But it is not immediately obvious why, if there are systematic differences, we should expect mock juries to show a leniency asymmetry but actual juries to show no or even the opposite asymmetry.

Of course, mock juries make hypothetical decisions and real juries produce real verdicts. Some studies have examined whether mock jurors reach systematically different judgments when they are led to believe that their verdicts will have real consequences, but the results are quite mixed, with one experiment finding greater leniency, one finding less leniency, and one finding no difference (see Kerr & Bray, 2005). MacCoun (2005) considers relevant psychological theories of judgment and motivation and argues that they fail to predict any specific directional bias toward either leniency or severity in mock cases; e.g., real juries must confront the possibility that an innocent defendant could be convicted, but they must equally confront the real possibility that a guilty defendant will go free.

Most of the 11 mock jury studies in the MacCoun and Kerr (1988) meta-analysis used student mock juries, in experiments conducted in the 1970s and 1980s. This raises the possibility that those mock jury studies differ from studies of actual jury verdicts because of age, period, and/or cohort effects. We think this is unlikely, though not impossible. First, MacCoun and Kerr (1988) directly compared student and community samples in their Experiment 1; both samples showed a significant leniency asymmetry, although it was a bit stronger in the student sample, perhaps because the non-student jurors were less likely to reach a unanimous verdict in the brief allotted deliberation time. Second, Bornstein (1999) examined 26 different experimental mock juror comparisons of student and nonstudent adult samples; 21 of these failed to detect a main effect for student status, and 24 of them failed to detect an interaction of student status and other experimental treatments.Third, as seen in Figure 4, the punitive attitudes of young people and the college educated are only subtly different from those of older and/or less educated adults (Sourcebook of Criminal Justice Statistics Online, 2009a, 2009b).

And if anything, the trend since the 1970s has been toward greater leniency, not greater severity.

Mock jury experiments are not inherently superior to nonexperimental field studies, nor are they inherently inferior. It isn't even the case that mock jury experiments are inherently inferior to actual field experiments on jury behavior; the latter are scarce and difficult to conduct, they rarely permit unconfounded experimental manipulations, and they usually lack the statistical power of mock jury experiments. Real trials inevitably provide a noisier sample than mock jury studies, because there is only one jury per case in real trials, but often dozens of juries per case in mock jury studies. Ceteris paribus, this noise should make it more difficult to isolate specific empirical regularities like the leniency asymmetry. In contrast, to examine the leniency asymmetry, MacCoun and Kerr (1988, Experiment 2) were able to assemble 89 different mock juries, all considering the same criminal trial and starting deliberation with an even 2:2 split, so conditions were far more optimal for testing for an asymmetry effect

Should we always expect a leniency asymmetry?

Empirical research and theoretical modeling suggest that, in general, any juror/jury bias—whether proscribed or prescribed—is most likely to emerge in jury deliberations when cases are close (i.e., prosecution and defense cases are roughly equal in strength; cf. Kalven & Zeisel, 1966; Kerr et al., 1996; Kerr, Neidermeier, & Kaplan, 1999). Thus there is no reason to expect the leniency asymmetry to operate as some sort of universal constant (like the speed of light); rather, it should be seen as one parameter of group process that will varies across cases and settings. In our 1988 paper (p. 31), we documented this point:

> …if the evidence for a particular verdict is clear (i.e., everyone agrees on the relevant conceptual system and facts), then the favored verdict should be demonstrably "correct" and a "bias" for that position should be evident in the social decision scheme.This analysis suggests that the stronger and clearer the evidence against the defendant is, the less pronounced the general leniency bias should be.To test this conjecture, we computed the correlation between the initial whole-sample predeliberation conviction rates and the asymmetry effect [viz., p (acquit | a verdict reached and an even initial split)] for the studies in Table 1.The two were negatively related, r (N = 10) = -.56, t (8) = 1.91, $p < .05$, one-tailed. The one nonlenient asymmetry effect becomes less puzzling; it was the case that had the strongest prosecution evidence. Thus, the asymmetry effect depends not only on the formal values imposed by the Court through its instructions to the jury but also on the overall weight of evidence against the defendant.

Unfortunately, we have no direct way of assessing the "initial whole-sample predeliberation conviction rates" for any of the actual jury datasets in Table 1. But we get some indication from the juries' initial splits, which suggest that most of the cases in the jury field studies had relatively strong evidence against the defendant. The ratio of the number of juries with a pro-conviction majorities to those with pro-acquittal majorities was 2.2 (= 148/67) for Kalven and Zeisel (1966), 3.2 (= 160/49) for Sandys and Dillehay (1995), 2.4 (= 68/28) for Devine et al. (2007); and 1.2 (= 140/116) for Hannaford-Agor et al. (2002); only for Devine et al. (2004) was this ratio below 1 (.89 = 33/37). Since the decision to go to trial for actual cases rests largely with prosecutors and grand juries, this makes sense—they would not generally risk the resources required to prosecute a case unless they thought there was a clear preponderance of evidence against defendants. By contrast, the goal in many jury simulation studies is to explore an empirical relationship of interest between some aspect of the crime, the defendant, the jury, or the trial procedure and some aspect of juror/jury behavior. As noted previously, such relationships tend to be attenuated for "open and shut" (i.e., non-close) cases, and thus, the stimulus trials used in jury simulation studies are usually relatively close (by design). For example, the weighted average of predeliberation conviction rates in the 11 simulation studies meta-analysed by MacCoun and Kerr (1988) was 53.6%. All this suggests that leniency asymmetry effect should be stronger in typical mock jury studies than in these studies of actual juries, not due to any lack of realism in their behavior, but because the mock jury studies used fairly close cases whereas these real-trial samples involve relatively pro-conviction cases. However, viewed in this light, this is less an issue of ecological validity (whether actual juries show a leniency asymmetry) than of the effect of moderating variables (i.e., the magnitude of the asymmetry will depend on the strength of evidence against the defendant for any type of jury).

What would the absense of a leniency asymmetry imply about actual trials?

Oddly, neither Devine and his colleagues (2004, 2007) nor Sandys and Dillehay (1995) make any reference to what we see as the principle contribution of our 1988 paper—our test of a theoretical explanation for this phenomenon. Specifically, we argued that the operative standard of proof in criminal trials—the reasonable doubt standard––might provide certain advantages to advocates of acquittal. One is a rhetorical advantage; it should be easier to raise a single reasonable doubt than to refute all possible reasonable doubts. Another is the implication of a reasonable doubt standard for the jury's decision rule—the defense, not the prosecution, should get the advantage of any doubts (i.e., "ties" should go to the defense). And finally, from a social comparison standpoint, the very fact of a sharp disagreement among reasonable people acting in good faith might be taken as evidence of reasonable doubt.

Results from two previous experiments (Kerr et al., 1976; MacCoun, 1984) that manipulated standard of proof instructions were consistent with this explanation, but these studies had too little power in the critical 50:50 condition to provide a strong test. So in a new experiment (MacCoun & Kerr, 1988, Exp. 2) we tested the effect of the reasonable doubt standard vs. the symmetrical "preponderance of evidence" standard, on mock juries that were composed (using predeliberation questionnaire data) of two advocates for acquittal and two advocates for conviction. As noted earlier, we found a significant leniency asymmetry in the reasonable doubt condition, but no asymmetry in the preponderance of evidence condition.

In this paper, we have shown that after eliminating biases in coding, actual juries tend to show a leniency asymmetry (albeit somewhat attenuated) that has been well documented in mock juries. Although we have not focused on them here, there are other explanations for mock vs. real jury differences besides the how splits are coded. For example, the field studies rely on jurors' fallible recollections of initial splits rather than on predeliberation responses. And jurors may be more likely to vote on the first public ballot contrary to their actual pre-deliberation preference in the direction of acquittal vs. conviction (see Note 4, *Supra*). We must await new research to determine how useful such explanations may be. However, at this point, it seems fair to conclude that the case against a leniency asymmetry in actual juries is far weaker than it may have appeared to be.

But for the sake of argument, let us suppose that newer and stronger evidence emerges for a symmetrical pattern of influence in real juries—or even an asymmetry favoring severity. We would view such a pattern not as an indictment of the mock jury method, but rather as an indictment of actual jury behavior. True, actual jury trials are what we use to determine the guilt of actual defendants. But that is a descriptive fact, not a normative argument. Perhaps the proper framing is not "real vs. fake," but—to adopt a useful psycholinguistics distinction—"performance vs. competence." The mock jury studies show that citizens are capable of translating the reasonable doubt instructions into a rhetorical advantage for advocates of acquittal. The question is, do they actually make use of this competence when they perform in real trials? If actual juries fail to show any leniency asymmetry, perhaps rather than asking "what's wrong with the mock jury method?" we should instead be asking "are actual juries failing to give criminal defendants the benefit of a reasonable doubt?" Clearly, we need more and better evidence on the function relating initial splits to final verdicts for actual criminal juries (i.e., their social decision scheme) before we can seriously begin to debate such issues.

Postscript

This special issue of GPIR is a tribute to the work and career of James H. Davis. Potential contributors were exhorted not to eulogize Jim or his many contributions to the field, but told that " … rather than looking backward at Jim's work, we are asking contributors to present their own work or perspective—empirical, theoretical, or conceptual—that they feel would move the field forward." In this paper, we have stuck to this model. However, in closing, we would like briefly to make a few points. It was Jim Davis' flexible and powerful social decision scheme (SDS) model (Davis, 1973; Stasser, 1999) that provided a conceptual and methodological framework for exploring and documenting asymmetries in the jury decision making process. More generally, such asymmetries have proven to be highly diagnostic of interesting patterns in group decision making and problem solving (e.g., Kerr et al., 1996; Laughlin & Ellis, 1986; Tindale, Smith, Thomas, Filkins, & Sheffey, 1996). It was Jim Davis who, to our knowledge, first speculated about the possibility of a leniency asymmetry in juries (e.g., Davis, 1973; Davis, Kerr, Atkin, Holt, & Meek, 1975); he referred to it as a *defendant protection norm*. Further, his early jury experiments provided an empirical foundation for not only this phenomenon, but many other interesting jury phenomena (see Stasser, Kerr, & Davis, 1989; Tindale, Nadler, Krebel, & Davis, 2003 for a partial review). In short, the present paper could not have been conceived, let alone written, without the many conceptual, methodological, and empirical contributions Jim Davis made to the study of group behavior.

Notes

We wish to thank Dennis Devine, Ron Dillehay, and Paula Hannaford-Agor for providing data from their field studies, and Dan Simon for helpful comments. Portions of this paper were presented at the 2011 annual meeting of the Midwestern Psychological Association, and at the 2011 meeting of the Society for Experimental Social Psychology.

1 It might be noted that there remained disagreement even in these 79 juries. If it existed, a modal judgment or the median of contiguous first-ballot estimates was taken to be sufficient agreement for purposes of coding. Hence, the ostensive agreement rate of 68% must be viewed as an upper bound on the level of agreement between jurors.

2 The most significant difference is that rather than fill out questionnaires immediately after the trial, jurors were given questionnaires and asked to fill them out and mail them back later.

3 Sandys and Dillehay classify juries with respect to the % voting guilty on first ballots, and elsewhere in their paper (p. 188, 1st paragraph) note that some initial-ballot votes were undecideds. Kalven and Zeizel (1966) explicitly note that undecideds were combined with Not Guilty votes (footnote to Table 138, p. 487).

4 It is also conceivable that first-ballot Not Guilty votes overstate predeliberation support for acquitting the defendant. Jurors are explicitly instructed to presume innocence until there is proof beyond a reasonable doubt, and uncertain or cautious jurors might endorse "Not Guilty" early in deliberation out of a sense of procedural fairness. Sandys and Dillehay (1995, p. 192) made a similar observation: "Complex or unclear evidence may lead to a cautious initial vote (a leniency effect), which is then overcome by a careful review of the evidence. " The net effect of such behavior would, like the coding assumption discussed in the main text, be to misclassify juries as more pro-acquittal than they would be had predeliberation verdict preferences been available. In the only study explicitly examining this issue, Davis et al. (1988) found that mock criminal jurors were significantly more likely to shift from a predeliberation vote for Guilty to a first ballot vote for Not Guilty than in the opposite way. Although it may be unavoidable in field studies of jury behavior, reliance on first ballot (rather than predeliberation) preferences therefore seems

likely to introduce yet another bias into the classification of jury initial splits.

5 Certain conventions had to be adopted in the recoding and re-analysis of each of these studies. For example, in Devine et al. (2007), one jury that voted two Guilty and ten Not Guilty acquitted on the original charge but convicted on a lesser included charge. Devine et al. (2007) classified this as a Guilty verdict (see footnote to Table 3). To be consistent with the remaining cases (where a failure to convict on the initial charge was counted as an acquittal), we counted this jury as an acquittal (hence, the row for 17% G (= 2/12) in Table 4 has seven acquittals and no convictions, while the corresponding row in Table 3 has six acquittals and one conviction).

6 Disagreement among jurors about the initial ballot split on the primary charge was common in these data. For example, for only five of the 51 focal trials did all respondents recall the first-ballot identically; these were all unanimous juries. For seven trials, the level of agreement was so low that a reliable estimate of first-ballot split simply could not be made. For the remaining 44 trials, following Sandys and Devine's (1995) procedure, if 2/3 or more of the respondents did agree on a split, that split was coded; otherwise, the average split across respondents was used.

7 Many of the trials in this data set involved multiple charges or defendants. It was generally not possible in such trials to unambiguously determine the verdict for the first charge, for which the initial-ballot split was obtained. Therefore, in this reanalysis, we restricted our attention to trials with a single charge and defendant.

8 In Figures 2 and 3 and in the text's discussion of the figures, the percentage rather than the proportion of guilty votes at the start of deliberation are plotted, and the b parameter is also expressed in percentage terms. This is done simply for ease of presentation. In the actual calculations, and in the summaries at the bottom of Table 4, proportions, not percentages are always used.

References

Bornstein, B. H. (1999).The ecological validity of jury simulations: Is the jury still out? *Law and Human Behavior, 23*, 75–91.

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

Davis, J. H., Kerr, N. L., Atkin, R., Holt, R., & Meek, D. (1975). The decision processes of 6- and 12-person mock juries assigned unanimous and 2/3 majority rules. *Journal of Personality and Social Psychology, 32*, 1–14.

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

Davis, J. H., Stasser, G., Spitzer, C. E., & Holt, R. W. (1976). Changes in group members' decision preferences during discussion: An illustration with mock juries. *Journal of Personality and Social Psychology, 34*, 1177–1187.

Devine, D. J., Buddenbaum, J., Houp, S., Studebaker, N., & Stolle, D. P. (2007). Deliberation quality: A preliminary examination in real juries. *Journal of Empirical Legal Studies, 4*, 273–303.

Devine, D. J., Olafson, K. M., Jarvis, L. L., Bott, J. P., Clayton, L. D., & Wolfe, J. M. T. (2004). Explaining jury verdicts: Is leniency bias for real? *Journal of Applied Social Psychology, 34*(10), 2069–2098.

Devine, D. J., Clayton, L. D., Dunford, B. B., Seying, R., & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy, and Law, 7*(3), 622–727.

Hannaford-Agor, P. L., Hans, V. P., Mott, N. L., & Munsterman, G. T. (2002). Are Hung Juries a Problem? Final Report, National Center for State Court, to National Institute of Justice, September 30, 2002.

Hans, V. P., Hannaford-Agor, P. L., Mott, N. L., & Munsterman, G. T. (2003). The hung jury: *The American Jury*'s insights and contemporary understanding. *Criminal Law Bulletin, 39*, 33–50.

Kalven, H., & Zeisel, H. (1966). *The American Jury*. Boston, MA: Little-Brown.

Kerr, N. L., Atkin, R. S., Stasser, G., Meek, D., Holt, R. W., & Davis, J. H. (1976). Guilt beyond a reasonable doubt: Effects of concept definition and assigned decision rule on the judgments of mock jurors. *Journal of Personality and Social Psychology, 34*, 282–294.

Kerr, N. L., & Bray, R. M. (2005). Simulation, realism, and the study of the jury. In N. Brewer, & K. D. Williams (Eds.), *Psychology and law: An empirical perspective*. New York, NY: Guilford.

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

Kerr, N. L., & MacCoun, R. J. (2007). Juries and the leniency bias. In B. Cutler (Ed.), *The Encyclopedia of Psychology and Law*. Thousand Oaks, CA: Sage.

Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996) Bias in judgment: Comparing individuals and groups. *Psychological Review*, 103, 687–719.

Kerr, N. L., Neidermeier, K. E., & Kaplan, M. F. (1999). Bias in jurors vs bias in juries: New evidence from the SDS perspective. *Organizational Behavior and Human Decision Processes, 80*, 70–86.

Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellective tasks. *Journal of Experimental Social Psychology, 22*, 177–189.

MacCoun, R. J. (1984). Modeling the impact of extra-legal bias and defined standards of proof on the decisions of mock jurors and juries. *Dissertation Abstracts International, 46*, 700B.

MacCoun, R. J. (2005). Comparing legal factfinders: Real and mock, amateur and professional. *Florida State University Law Review, 32*, 511–518 .

MacCoun, R. J. (in press). The burden on social proof: Shared thresholds and social influence. *Psychological Review.*

MacCoun, R. J., & Kerr, N. L. (1988).Asymmetric influence in mock jury deliberation: Jurors' bias for leniency. *Journal of Personality and Social Psychology*, *54*, 21–33.

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

Sourcebook of Criminal Justice Statistics Online (2009a). Table 2.50: Attitudes toward the death penalty for persons convicted of murder, by demographic characteristics, United States, selected years 1980–2002. Retrieved August 10, 2009, from http://www.albany.edu/sourcebook/csv/t250.csv

Sourcebook of Criminal Justice Statistics Online (2009b). Table 2.47: Attitudes toward severity of courts in own area, by demographic characteristics, United States, selected years 1985–2002. Retrieved August 10, 2009, from http://www.albany.edu/sourcebook/csv/t247.csv

Stasser, G. (1999). A primer of social decision scheme theory: Models of group influence, competitive model-testing, and prospective modeling. *Organizational Behavior and Human Decision Processes, 80*, 3–20.

Stasser, G., Kerr, N. L., & Bray, R. M. (1982).The social psychology of jury deliberations: Structure, process, and product. In N. L. Kerr, & R. M. Bray (Eds.), *The psychology of the courtroom*. New York, NY: Academic Press.

Stasser, G., Kerr, N. L., & Davis, J. H. (1989). Influence processes and consensus models in decision-making groups.In P. Paulus (Ed.), *Psychology of group influence* (2nd Ed.). Hillsdale, NJ: L. Erlbaum.

Stawiski, S., Dykema-Engblade, A., & Tindale, R. S. (2012). The roles of shared stereotypes and shared processing goals on mock jury decision making. *Basic and Applied Social Psychology, 34*, 88–97.

Tindale, R. S., Nadler, J., Krebel, A., & Davis, J. H. (2003). Procedural mechanisms and jury behavior. In M. A. Hogg & R. S. Tindale (Eds.), *Blackwell Handbook of Social Psychology: Group Processes*. Oxford: Blackwell.

Tindale, R. S., Smith, C. M., Thomas, L. S., Filkins, J., & Sheffey, S. (1996). Shared representations and asymmetric social influence processes in small groups. In E. H. Witte, & J. H. Davis (Eds.), *Understanding group behavior: Consensual action by small groups* (Vol. 1) (pp. 81–103). Hillsdale, NJ: Lawrence Erlbaum Associates.

Williams v. Florida, 399 US 78 (1970).

Zeisel, H., & Diamond, S. S. (1978). The effects of peremptory challenges on jury and verdict: An experiment in a federal district court. *Stanford Law Review, 30*, 491–531.