Law Hum Behav (2009) 33:481–496
DOI 10.1007/s10979-008-9168-2

ORIGINAL ARTICLE

Capital Jury Deliberation: Effects on Death Sentencing, Comprehension, and Discrimination

Mona Lynch · Craig Haney

Published online: 31 March 2009

Abstract This study focused on whether and how deliberations affected the comprehension of capital penalty phase jury instructions and patterns of racially discriminatory death sentencing. Jury-eligible subjects were randomly assigned to view one of four versions of a simulated capital penalty trial in which the race of defendant (Black or White) and the race of victim (Black or White) were varied orthogonally. The participants provided their initial "straw" sentencing verdicts individually and then deliberated in simulated 4–7 person "juries." Results indicated that deliberation created a punitive rather than lenient shift in the jurors' death sentencing behavior, failed to improve characteristically poor instructional comprehension, did not reduce the tendency for jurors to misuse penalty phase evidence (especially, mitigation), and exacerbated the tendency among White mock jurors to sentence Black defendants to death more often than White defendants.

A majority of the United States Supreme Court appears firmly entrenched in the view that capital punishment is not per se unconstitutional. However, several of the Court's decisions have insured that the behavior of capital juries will continue to hold center stage in the legal, political, and public debates over the death penalty, at least for the foreseeable future. Specifically, in *Ring v. Arizona* (2002), the Court underscored the importance of the jury's role in capital cases and ruled that "[c]apital defendants… are entitled to a jury determination of any fact on which the legislature conditions an increase in their maximum punishment" (p. 589). In addition, the Court recently reaffirmed its long-standing commitment to allowing capital juries to consider a broad range of mitigating evidence—literally "any aspect of the defendant's character or record and any of the circumstances of the offense that the defendant proffers as a basis for a sentence less than death" (*Lockett v. Ohio*, 1978, p. 604)—and overturned cases in which the sentencing instructions prevented capital juries from giving "meaningful consideration and effect to all mitigating evidence that might provide a basis for refusing to impose the death penalty" (*Abdul-Kabir v. Quarterman*, 2007, p. 1664).

Of course, it was the "unbridled discretion" that juries historically had been afforded in death penalty cases that led to the landmark decision in *Furman v. Georgia* (1972), declaring the nation's system of death sentencing unconstitutional because of the arbitrary and capricious outcomes that it produced. State legislatures reacted to *Furman* by, among other things, revising their capital sentencing statutes in ways that were supposed to "guide" juries in the exercise of that discretion. Yet, even now, as the current Court emphasizes the significance of the capital jury's factual determinations and insists that jurors be provided with all possibly relevant mitigating evidence (e.g., *Rompilla v. Beard*, 2005; *Wiggins v. Smith*, 2003), there is much evidence to suggest that many of the concerns raised in *Furman* about the apparent unfairness and unreliability of capital jury verdicts have still not been effectively addressed.

Indeed, more than 30 years have passed since the Court re-instituted capital punishment by approving a more guided and structured death sentencing process that presumably narrowed the capital jury's discretion and

M. Lynch
University of California, Irvine, Irvine, CA, USA

C. Haney (✉)
University of California, Santa Cruz, Santa Cruz, CA, USA
e-mail: psylaw@ucsc.edu

regularized its decision-making (*Gregg v. Georgia*, 1976). Yet, legal analysts and social scientists continue to warn that many of the pre-*Furman* problems persist. As one commentator concluded, "[t]he best available evidence shows that arbitrariness still runs rampant more than three decades after *Furman*" (McCord, 2005, p. 806) and that, as a result, "much of [the Court's] capital jurisprudence over the last three decades has been a colossal mistake" (p. 868). Another recently argued that "the best description of the capital sentencing pattern the Court's doctrine currently requires… is virtually identical to the pattern *Furman* ruled unconstitutional" (Liebman, 2007, p. 12).

Much social science research examining the nation's system of death sentencing has focused on two important flaws in the capital jury decision-making process. The first is its unreliability (i.e., the apparent arbitrariness and unpredictability in the outcomes of capital cases and the manner in which jurors reach them). These studies have suggested that although the capital jury's decision to take or spare a defendant's life is supposed to reflect the "reasoned moral response" of its members (*California v. Brown*, 1987, p. 545), aspects of the jury decision-making process are compromised by confusion and misunderstanding, leading to what Bowers and Steiner (1999) have termed "false and forced choices" (p. 605). The second flaw in the capital jury decision-making process is unfairness (i.e., the extent to which extralegal factors—especially race—influence the verdicts that are rendered in an appreciable number of death penalty cases). Thus, data suggest that capital jurors may actually dispense *less* justice than their non-capital counterparts, perhaps because jurors selected to serve in capital cases are especially susceptible to extralegal factors or influences (e.g., Butler, 2007). In any event, there is reason to believe that one particularly pernicious such influence—the consideration of race—continues to play a significant role determining which defendants are sentenced to die and which to live (e.g., Baldus, Woodworth, Zuckerman, Weiner, & Broffitt, 1998).

Our own research has focused on the interrelationship between these two serious flaws—unreliability and unfairness—in the capital jury decision-making process. Specifically, we have examined whether and how capital jurors' lack of instructional comprehension overall (Haney & Lynch 1994, 1997) contributes not only to confused decision-making but also to racially discriminatory death sentencing. Thus, in an earlier study (Lynch & Haney, 2000), we confirmed the fact that, much like the college student participants in our prior research, jury-eligible adults had great difficulty in comprehending the standard California death penalty instruction. A very high percentage was unable to correctly define the basic concepts of "aggravation" (reasons to sentence a defendant to death) and, especially, "mitigation" (reasons to sentence a defendant to life), and could not accurately identify which specific factors enumerated in the instruction pertained to each concept.

The widespread instructional incomprehension of capital penalty phase instructions that we documented in California is consistent with the findings of other researchers who have found similar levels of confusion among jurors in other states (e.g., Diamond, 1993; Eisenberg & Wells, 1993; Wiener et al., 2004). However, we also found that the level of our participants' instructional comprehension was related to their discriminatory death sentencing. That is, persons who poorly comprehended the instructions tended to sentence African-American capital defendants to death at significantly higher rates than jurors with high levels of comprehension. In addition, specific pieces of mitigating evidence—factors that are supposed to be used as the basis for life rather than death sentences—were not weighted as heavily for Black as for White defendants, and they were more likely to be used inappropriately (i.e., as favoring death rather than life) when the defendant was Black.

Like most research on jury decision-making, however, our initial study was focused on the responses of individual participants rather than the behavior of persons interacting in a jury or jury-like setting. Of course, we recognized the potential significance of jury deliberations, a topic that has been much discussed in psycho-legal research. As one early researcher put it, "the effects of group, as opposed to individual, decision making in jury research is too important to be ignored" (Bernard, 1979, p. 111). Among other things, analysts have questioned whether jury deliberations serve generally to either worsen or to ameliorate individual juror biases (e.g., Kerr, MacCoun, & Kramer, 1996). Thus, some studies have found that deliberation has a salutary, tempering effect on jurors' biased decision-making (e.g., Carretta & Moreland, 1983; Kerwin & Shaffer, 1994; London & Nunez, 2000), although others have documented some of the ways that it can exacerbate certain pre-existing problematic tendencies (e.g., Cowen, Thompson, & Ellsworth, 1984; Kramer, Kerr, & Carroll, 1990). Moreover, a number of researchers have examined whether deliberations lead to polarization on juries (whereby initial differences become more extreme), or produce the so-called "leniency shifts" in which individual jurors are generally more favorably disposed toward defendants after they have had an opportunity to deliberate (e.g., MacCoun & Kerr, 1988; Thompson, Fong, & Rosenhan, 1981).

For the most part, experimental research has examined the effects of deliberation by measuring rates of conviction. Few, if any, laboratory studies have focused on the effects of deliberation on death sentencing. Yet, there are a number of reasons to believe that the opportunity to

deliberate might have an especially significant impact on capital decision-making. For one, the decision of whether to sentence someone to die is inherently subjective and, as such, seems particularly susceptible to group-based norms that may emerge in the course of deliberation. Moreover, levels of instructional comprehension and tendencies to discriminate in death sentencing might be heavily influenced by the presence of others. For example, it is possible that a jury whose individual members did not necessarily all understand the capital sentencing instructions particularly well would draw on the collective wisdom of the group to elevate their overall comprehension. Similarly, the need to publicly share and discuss sentencing rationales with others might neutralize or dampen the expression of jurors' racial prejudices and lead to more even-handed decision-making.

With these things in mind, the present study was designed to explore whether and how deliberations affected the death-sentencing process, especially as they might influence instructional comprehension and jurors' reactions to the racial dynamics of the case.

METHOD

Participants and Study Site

Participants were recruited from a large, urban county in California. To insure that they were representative of actual California capital jurors, all participants were required to be both jury-eligible and death-qualified. To obtain an adequate pool of persons who met these criteria, we used a multi-pronged recruitment effort that included distributing flyers, posting notices on local electronic bulletin boards, and placing classified ads in regional newspapers that solicited adults to serve as mock jurors for a research project. Potential participants were promised $40.00 for approximately 3–4 h of their time.

Interested persons were given a local telephone number to call, which allowed a research assistant to pre-screen them for jury eligibility (i.e., to determine that they were at least 18 years old, had no felony convictions, were a United States citizen, and spoke sufficient English to understand the court proceedings) and also to death-qualify them. To determine whether potential participants met the standard death qualification criteria, as articulated in *Witherspoon v. Illinois* (1968), *Wainwright v. Witt* (1985), and *Morgan v. Illinois* (1992) each one was asked a series of questions about his or her death penalty attitudes. Persons who were unalterably opposed to or in favor of the death penalty in all cases, or acknowledged that the strength of their beliefs about the death penalty would interfere with or impair their ability to act as fair and impartial jurors, were excluded from participation. Only those who were jury-eligible *and* death-qualified were scheduled for participation.

To maximize the participation of local community residents, we rented office space in a central, accessible location. We also scheduled research sessions over a broad range of time periods, including weekends, to allow persons who were employed full-time to participate. A diverse group of eligible participants was obtained over the eight-month data collection period. We created a trial viewing and deliberation room inside the rented office space by equipping it with a television/VCR viewing unit, and a conference table and chairs. A video camera was installed to videotape deliberations.

Experimental Procedure

A total of 539 people participated in the study as members of 100 mock juries. We attempted to schedule seven participants for each jury session. Whenever fewer participants could be scheduled or actually appeared for each session, we required a minimum of four participants to constitute a "jury." If fewer than four people attended, the session was canceled, participants in the canceled sessions were paid for coming, and they were offered a chance to reschedule.

After at least four participants arrived at the study site, they were informed of their rights as research participants, and asked to sign a standard human subjects consent form. They then were given a short questionnaire that assessed their individual attitudes about a range of criminal justice issues, including the death penalty. Participants were told that the defendant in the case that they would be asked to judge had already been found guilty of first degree murder with special circumstances, and that their task as jurors would be to reach a sentencing verdict. Thus, just as in our previous study, we employed a penalty phase-only design. We did so primarily to simplify the task at hand, and to allow the participants to focus exclusively on penalty phase decision-making process. We reasoned that, although this meant that the simulation differed in some respects from the typical capital jury deliberation, it would, if anything, improve the jurors' performance by reducing the amount of guilt-phase "noise" that might otherwise complicate their decision-making. It also allowed us to focus more clearly on aspects of the death-sentencing process.

Jurors were told that the videotaped trial that they would watch—and on which they were to base their sentencing decision—was taken from proceedings that occurred in an actual case. Each participant was supplied with a pad and pen and encouraged to take notes to aid in their decision-making. Jurors then watched, as a group, one of four videotapes of the penalty phase of a simulated capital trial. The videotaped simulation was same one that had been

used in our earlier study of capital juror decision-making (Lynch & Haney, 2000, pp. 342–343). It was designed to capture common features of a "typical" capital murder case that pre-testing had shown contained "mid-range" case facts—ones that were neither so aggravated nor so mitigated that participants would vote overwhelmingly for one or another sentencing verdict. The case facts themselves were based on—but substantially abbreviated from—an actual capital penalty trial that occurred in a robbery murder case. The prosecution's penalty phase aggravation consisted of the circumstances of the crime, victim impact testimony, evidence that the defendant had been accused of a prior similar crime, and his apparent lack of remorse for the present crime. Defense mitigation included evidence that the defendant had been seriously abused as a child, had suffered from alcohol and drug abuse at the time of the crime, had pre-existing psychiatric problems that may have impaired his mental state, and testimony that he was still loved and needed by his family.

Each videotape began with a very brief synopsis of the guilt-phase case, followed by prosecution and defense opening statements. The presentation of evidence proceeded as it would in an actual penalty trial (prosecution first, defense second), and included the cross-examination of all witnesses. The tapes concluded with final arguments being given by both the prosecution and defense, and the judge's reading of the standard patterned jury instructions that were in use in capital penalty trials at the time in California.[1] The four 55-min versions of the simulated penalty trial were identical in all respects *except* that the race of defendant and race of victim were systematically varied (by using different photographs and the voices of different actors that pre-testing indicated were closely matched on the basis of physical attributes such as age, size, and appearance, and in performance style, including intonation and skill level). This resulted in the creation of four separate experimental conditions: Condition 1—White defendant/White victim; Condition 2—Black defendant/Black victim; Condition 3—White defendant/Black victim; and Condition 4—Black defendant/White victim.

After jurors finished viewing the videotape, but before they had an opportunity to discuss the case, they were asked to confidentially and individually make a written verdict choice and to indicate their confidence in that choice. Participants were told that this was a non-binding "straw vote" that they were free to change during or after deliberation. Each group of jurors was asked to select a foreperson and then to deliberate to reach a unanimous sentencing verdict. The sentencing verdict was a dichotomous choice of either life without the possibility of parole or the death penalty, and was recorded on a written verdict form. The jury was also asked to indicate its collective confidence in their decision on a seven-point scale. Each jury was given copies of the judicial instructions that had been read by the judge in the trial tape they had watched, and a verdict form on which to record the sentencing decision. The video camera was turned on to record deliberations, and the jurors were left alone in the jury room to deliberate.

The deliberation period was limited to a maximum of one and a half hours. If the jury was unable to reach a unanimous verdict within that time limit, it was declared deadlocked, and the nature of the split was noted on a "mistrial voting form" (including each individual juror's verdict and confidence rating). When the deliberation was completed and the instructions and the verdict forms had been collected, each individual juror was asked to complete a series of questionnaires. Jurors also completed a "juror perceptions" questionnaire that contained several open-ended questions about the decision-making process as well as a series of forced choice questions that asked them for ratings of the defendant, attorneys, witnesses, penalty trial evidence, and arguments presented. In addition, participants completed an instructional comprehension questionnaire (identical to the one used in Lynch & Haney, 2000), and a demographics questionnaire that included five items from the subtle racism scale developed by Pettigrew and Meertens (1995), which were the final items completed by our participants. The subtle racism measures assess attitudes about such issues as the degree to which minorities deserve equal opportunities, benefit from affirmative action, and suffer structural inequality, framed in language that is designed to tap contemporary forms of racism (see Pettigrew and Meertens, 1995, for more details).

Once they had completed the final set of questionnaires, participants were paid $40.00 in cash, debriefed, and asked not to talk to others in the community about the case or the study for the next few months (until data collection had ended).

RESULTS

We were primarily interested in examining the influence of group deliberation on the operation of several variables that our earlier research had determined were important to the capital sentencing process—namely, instructional comprehension and the racial dimensions of the case. Accordingly, the group-level data—the unanimous verdicts and verdict splits that were recorded in each condition—are the first that we present after describing the participants'

[1] These are the California pattern instructions, CALJIC 8.85 and 8.88, which were used in capital penalty trials in California for many years preceding and at the time the study was conducted.

general characteristics. We next discuss deliberation-related changes (i.e., how individual juror responses were influenced and affected by the opportunity to engage in jury-level discussion and interaction, and the need to seek a group consensus on a sentencing verdict). This section presents results that address the differences between those participants who voted for life compared with those who opted for a death sentence. We then separately examine how the racial dynamics of the different case scenarios and the participants' level of instructional comprehension affected their pre- and post-deliberation decision-making. Because this study was in many ways a replication of our previous research (Lynch & Haney, 2000), a number of the current findings are compared to the results of our earlier study.

Participant Demographics

The group of 539 participants was composed of 61% women and 39% men, ranging in age from 18 to 84 years old, with a mean age of approximately 48 years. As would be expected from the racial composition of the county in which the study took place, and the tendency of death qualification to disproportionately eliminate racial minorities, participants overwhelmingly identified themselves as White (82%). A total of 8% identified as Latino/a, 4% as Asian-American, and 4% were African-American. They were relatively well educated overall: the great majority (81%) had at least some college education, and almost half of those had at least 4-year degree.

Over half of our participants (52%) reported some kind of current employment, and another 25% of them were retired. The remainder identified themselves as either unemployed or full-time students. About half of our participants (51%) were self-described Democrats, 21% identified as Republican, and the remainder indicated either some other political affiliation or none at all. In terms of political ideology, 25% considered themselves conservative, 45% identified as moderate, and 28% indicated that they were liberal. Just over one-third (34%) of the participants had actually served on a jury at some time before participating in this study.

As also would be expected, given the fact that all of the 539 participants were death-qualified prior to participating, three quarters (76%) were somewhat or strongly in favor of the death penalty, 20% identified themselves as somewhat opposed, and only 4% reported that they were strongly opposed.

Voting for Life or Death: Group Outcomes

A total of 61 of our 100 deliberating juries reached unanimous verdicts. Of those, the majority (45 of 61, 74%) voted for a death sentence. Among the remaining 39 juries that were unable to come to a unanimous decision, 20 had majorities that favored death sentences, 13 had majorities that favored life without parole, and 6 were evenly split between life and death. Thus, taking into account only those juries that had unanimous verdicts or majority votes ($N = 94$), 69% were in favor of death, and 31% supported a verdict of life without possibility of parole.

Juries were non-significantly more likely to prefer a sentence of death when they were considering the Black defendant cases. In these conditions, 77% of juries leaned toward or unanimously decided on death, whereas in the White defendant conditions, only 62% leaned toward or unanimously chose death (Pearson Chi-square = 2.44; $p < .12$, 2-tailed; Cramer's $V = .16$). This differential was more pronounced when comparing only the two cross-racial conditions: excluding the cases where the jury was evenly split, 79% of the 24 juries who viewed the Black defendant/White victim trial tape (Condition 4) leaned toward or unanimously voted for death, but only 56% of the 23 juries in the White defendant/Black victim condition (Condition 3) favored death. This difference was marginally significant (Chi-square = 2.77; $p < .10$; df = 1; Cramer's $V = .24$). We should note that based on our power analysis of expected effect size, we assumed that we would need approximately 400 total juries to uncover main race effects at the group level, so we did not expect that these differences would reach statistical significance.

The other predictor of death verdicts or majority splits in favor of death at the group level was the ratio of White participants on the jury. Specifically, the higher the proportion of Whites on the jury, the more likely the jury was to favor death ($B = 1.942$; $p < .05$; $R^2 = .05$). When we examined the jury groups separately depending upon whether they viewed the Black defendant or the White defendant, we found that the proportion of Whites on the jury was a significant predictor of death verdicts in the Black defendant condition only ($B = 3.142$, $p < .01$, $R^2 = .14$).

Predicting Life or Death Outcomes: Individual Straw Votes and Final Verdicts

Participants' straw votes indicated that a little more than half (54%) of them favored a death verdict after hearing the penalty phase evidence but before deliberating. After deliberation, however, there was a significant shift toward death, with 66% of all participants sentencing the defendant to die. A paired samples t-test indicated that this within subject shift toward death was significant for verdict certainty (t (535) = 5.71; $p < .0001$).

We also found some differences in verdict preference as a function of participant demographics. Thus, women were significantly less inclined to sentence the defendant to death than were men in their pre-deliberation straw votes: only 50% of the women chose death, as compared to 60% of the men who did (Chi-square = 4.57; d.f. = 1; $p < .05$; Cramer's $V = .09$). However, the gender difference largely disappeared after deliberation; 65% of the women and 68% of the men selected death as their final verdict choice. Thus, although women outnumbered men in our participant pool, it appeared that they actually were less influential in the deliberation process and had a somewhat greater tendency to shift toward death than did the men (who, of course, were still somewhat more likely to favor death overall).

In addition, we found that individual non-White participants were significantly less likely to render death sentences than were White participants. This was true at both the straw vote and final vote stages: 39% of the non-White jurors chose death in their straw votes, compared to 57% of the White jurors (Chi-square = 10.4; d.f. = 1; $p = .001$; Cramer's $V = .14$). Although non-Whites did shift toward death at a greater rate than Whites following deliberation, they were still significantly less likely to choose the death penalty as a final verdict. Specifically, the final verdict counts indicated that slightly more than half (54%) of non-White participants chose death as their final verdict, as compared to 69% of Whites who did. Hierarchical linear mixed models analysis indicates that the two groups significantly differed in this regard ($F(1,433) = 6.85, p < .01$). These data suggest that White men were the least likely to change their positions from the straw vote to the final vote, and appeared to disproportionately influence women and non-Whites in the deliberation process.

In addition to gender and race/ethnicity, several attitudinal measures were correlated with the pre-deliberation straw votes. These included the participants' death penalty attitudes as well as several criminal justice attitude and subtle racism measures. We used logistic regression to identify those combinations of participant characteristics that best predicted individual straw votes and final verdicts. Using a forward entry logistic regression (LR), we found that death penalty attitude (Beta = 1.14; $p < .001$), whether the participant would consider mercy for "even the worst criminal" (Beta = −.61; $p < .001$), and whether the participant was White (Beta = −.60; $p < .02$) were very strong predictors of sentencing outcome at the straw vote stage (R^2 for model = .27). In terms of final individual verdict, jurors who were White, more strongly in favor of the death penalty, who believed that "everyone has an equal chance to succeed in this country, if they are just willing to try hard enough" (one of our subtle racism measures), and those who believed that Blacks should be able overcome prejudice "without special treatment" were more likely to sentence the defendant to death following deliberation.[2]

The Use and Misuse of Aggravating and Mitigating Evidence

As in our previous study (Lynch & Haney, 2000), we found that jurors who voted for death sentences used the aggravating and mitigating evidence differently than those who rendered life verdicts. Not surprisingly, aggravating evidence was seen as weighing in favor of death by many more jurors who actually voted for a death sentence than by those who voted for life. As Table 1 illustrates, the converse was also true; that is, mitigating evidence was much more likely to be seen as weighing in favor of life by pro-life jurors than by pro-death jurors.

However, there were also a number of jurors who *misused* penalty phase evidence by converting aggravation or mitigation into its opposite. This was much more likely to happen with mitigation—which was more often converted to aggravation—than the reverse. Thus, a number of pro-death jurors (between 14 and 30%, depending on the specific mitigating factor) actually weighed mitigating evidence as favoring a death sentence. For example, we found that three out of ten (30%) pro-death jurors improperly used the mitigating evidence that the defendant was impaired as a result of substance abuse, interpreting it as aggravation (i.e., as a factor weighing in favor of death) instead. This was two and a half times the number of death voters who weighed that particular piece of mitigating evidence as something in favor of a life verdict (12%).

Other mitigation suffered a similar (albeit not as extreme) fate at the hands of many pro-death jurors. Thus, nearly one quarter (23%) of jurors who chose death as their final verdicts incorrectly used the mitigating evidence that the defendant had experienced severe child abuse by weighing it on the side of the death penalty rather than a life verdict. On the other hand, as is clear from Table 1, this testimony was regarded as very important by pro-life jurors; fully 82% indicated that it weighed in favor of a life verdict. Yet, only 32% of the pro-death jurors gave it similar mitigating weight (as compared to the 23% of them who incorrectly weighed it in favor of death).

The Role of Race in Individual Sentencing Decisions

We found no main or interaction effect for defendant or victim race in the pre-deliberation *straw* vote outcomes, or

[2] Here we conducted a hierarchical mixed models analysis, which indicated that each of these variables significantly predicted verdict at $p < .01$.

Table 1 Weighing of aggravation and mitigation by final verdict choice

Evidence	Life			Death		
	Weighed toward life	No weight in decision	Weighed toward death	Weighed toward life	No weight in decision	Weighed toward death
Mitigating evidence						
Childhood abuse	82% (150)	15% (27)	3% (5)	32% (115)	45% (160)	23% (80)
Substance abuse	60% (109)	29% (52)	11% (21)	12% (43)	58% (205)	30% (107)
Psych. problems	74% (134)	22% (40)	4% (7)	19% (69)	62% (221)	18% (65)
Loved by family	67% (121)	31% (57)	2% (4)	21% (74)	65% (231)	14% (51)
Aggravating evidence						
Murder facts	25% (44)	15% (27)	60% (108)	3% (9)	5% (19)	92% (328)
Victim impact	23% (41)	43% (75)	34% (62)	6% (22)	35% (125)	59% (210)
Prior crime	26% (46)	29% (53)	45% (81)	7% (24)	10% (35)	83% (297)
No remorse	20% (36)	41% (74)	40% (72)	5% (17)	8% (29)	87% (310)

Note: Using hierarchical linear modeling, the differences by verdict choice on each of these pieces of evidence was highly significant ($p < .0001$)

Life verdict $N = 182$

Death verdict $N = 357$

1–4 missing values per group

for the verdict certainty measure, which we used as a more sensitive measure of verdict preference. Although Black defendants overall were slightly more likely to be sentenced to death at the straw vote stage (57% received death sentences as opposed to 52% in the White defendant conditions), the difference was not statistically significant. *After* deliberation, however, individual final verdict and certainty measures differed as a function of the race of the defendant. Specifically, the Black defendant was ten percentage points more likely to receive a post-deliberation death sentence (71%) from individual participants than was the White defendant (61%). A linear mixed model hierarchical analysis of jurors' verdict certainty, nested within jury, indicates a marginally significant overall defendant race effect ($F(1,106) = 2.89$, $p < .10$).

Given our initial findings that gender and race of jury members were predictive of verdict judgments and shifts in verdicts, and given existing literature that indicates that both race and gender of juror may influence sentence outcomes in a way that interacts with defendant race (Bowers, Steiner, & Sandys, 2001), we conducted several analyses that specifically addressed these possibilities. First, we examined the outcomes for White jurors only, nested within jury. When only the White jurors were included ($N = 439$), the race of defendant had a more pronounced effect on verdict certainty scores; those who saw the Black defendant indicated more death certainty than were those who viewed the White defendant ($F(1,102) = 4.36$, $p < .05$).

We also explored was has been termed a "White male dominance" effect on capital jury deliberations (Bowers et al. 2001) by analyzing the subgroup of White men mock jurors ($N = 169$) separately from the subgroup of women and non-Whites ($N = 365$). The results strongly suggested that the race effects for the overall group were driven by the White men on the mock juries. The White male jurors' assessments of verdict and verdict certainty varied significantly by defendant race. Of those who judged the Black defendant, 83% opted for the death sentence, while only 60% who saw the White defendant chose death. A hierarchical linear modeling analysis, nesting juror within jury, indicates that the difference in verdict is highly significant ($t = 2.92$, $p < .01$). Similarly, a linear mixed model test of the effect of defendant race on verdict certainty was highly significant ($F(1,79) = 11.75$, $p = .001$). In contrast, verdicts of the women and non-White jurors were not measurably affected by the defendant's race. However, when we isolated the White jurors' straw vote measures, then further isolated the White men jurors' straw vote measures, we did not find a significant defendant race effect. This suggested that something occurred in the deliberation process to activate racism within our White (and White male) jurors.

As noted above, there was a significant overall shift toward death from straw vote to final verdict, and we were interested in assessing whether the relative shifts across racial conditions differed. A linear mixed model hierarchical analysis of jurors' verdict certainty shift, nested in jury, indicated that jurors' change in verdict certainty from initial

straw vote to their final post-deliberation verdict significantly varied by defendant race ($F(1,98) = 13.87$, $p < .05$). Specifically, participants in the Black defendant conditions had a mean shift toward death of .94 on an 8-point scale, while those in the White defendant conditions moved .26 toward death overall. Given that nearly one quarter of the participants actually shifted toward life, and over one-third did not change their certainty scores at all, the net change of nearly a full point in the Black defendant conditions toward death is notable. Also notable is the fact that the pre- to post-deliberation verdict certainty shift toward death was smallest (.09) in the condition in which a White defendant was convicted of killing a Black victim (Condition 3).

The race of *defendant* effects described above appear mainly to be driven largely by the way in which the mitigating evidence, rather than the aggravating evidence, was used by our mock jurors. As Table 2 illustrates, the race of defendant influenced the way the jurors used the mitigating evidence that was introduced in the case. A linear mixed model hierarchical analysis revealed that individual level means, nested in jury, significantly differed for two of the four pieces of mitigating evidence (childhood abuse, substance abuse) as a function of defendant race (see Table 2). We also examined whether the defendant's race influenced the relative weight given to the aggravating evidence. As in our previous study, defendant's race did not significantly influence the use of aggravation.

Race of *victim* also had an impact on the relative weight given to two categories of mitigating evidence—that the defendant suffered from psychiatric problems and that he had substance abuse issues. When the victim was White, jurors gave significantly less weight to these specific pieces of mitigation than when the victim was Black. Victim race had no impact on how any of the aggravation was weighed (see Table 2).

Again, we isolated the White men mock jurors to see how their assessments of the evidence were impacted by defendant race. As is indicated in Table 3, below, all of the

Table 2 Mean weights of mitigating evidence by race of defendant and race of victim

Mitigating evidence	Mean for *Black* defendant	Mean for *White* defendant	*t*-value
Childhood abuse	3.42	3.11	−2.27*
Psych. problems	3.63	3.41	−1.44 (n.s.)
Substance abuse	4.02	3.74	−2.02*
Loved by family	3.57	3.50	−.30 (n.s.)
Mitigating evidence	Mean for *Black* victim	Mean for *White* victim	*t*-value
Childhood abuse	3.24	3.30	.36 (n.s.)
Psych. problems	3.40	3.64	2.06*
Substance abuse	3.75	4.01	2.05*
Loved by family	3.46	3.60	.88 (n.s.)

Total $N = 533$

Scale: 1 = weighed heavily for life, 4 = did not weigh in decision at all, 7 = weighed heavily for death

* $p < .05$ using linear mixed models hierarchical analysis with juror race nested within jury

Table 3 Effect of race on evidence assessment and verdict—White male jurors only

	Black defendant	White defendant	*t*-value for fixed effect
Mitigating evidence			
Childhood abuse	3.74	2.86	−4.51**
Substance abuse	4.14	3.53	−3.30**
Psych. problems	3.75	3.36	−2.40*
Loved by family	3.89	3.44	−2.59**
Aggravating evidence			
Murder facts	6.25	5.66	−2.87**
Victim impact	4.82	4.49	-1.92^{+}
Prior crime	5.48	4.98	−2.10*
No remorse	5.85	5.00	−2.93**

$N = 169$; ** $p < .01$; * $p < .05$; + $p < .10$, using linear mixed model hierarchical analyses, with juror nested in jury. Higher number means weighed more heavily toward death; 4 equals no weight given at all

mitigating evidence, and three out of four of the pieces of aggravating evidence (the facts of the crime, defendant's prior crime, and his lack of remorse) were all significantly influenced by the race of the defendant. Additionally, these jurors' assessment of the victim impact testimony was significantly influenced by race of victim, such that they weighed it more toward death for the White victim ($F(1,166) = 4.39$, $p < .05$). In contrast, when the women and non-White jurors were analyzed as a separate group, there were no differences in the weigh that the evidence was weighed as a function of defendant or victim race.

Instructional Comprehension

Instructional comprehension was measured after jurors not only had heard the instructions read aloud but also had an occasion to apply them to the particular set of case facts in a group deliberation setting. As we noted earlier, we thought that deliberations might provide an opportunity for those jurors who comprehended the instructions more accurately to correct those who understood them less well, thereby elevating overall comprehension scores. There was no evidence that this occurred. The average "correctness" score on the comprehension measure (14.7 correct out of a possible 35 points, or an average of 42% accurate) actually was *lower* than the one achieved by the non-deliberating participants in our prior research. Just 2 of the 539 participants scored 30 or above (86% correct)—the same number who gave absolutely no correct responses. Replicating our previous findings, we found that death penalty attitude was significantly and negatively correlated with comprehension score, so that jurors who expressed greater overall support for capital punishment tended to have lower instructional comprehension scores ($r = .12$, $p < .01$). Not surprisingly, instructional comprehension also was positively correlated with education level ($r = .26$, $p < .01$).

As in our previous study, we created a dichotomous variable for comprehension level, by splitting the group into two: those with a score of 14 or above were in the "high comprehension" group ($N = 288$) and those who scored 13 or below were in the "low comprehension" group ($N = 251$). Comprehension level was significantly related both to jurors' pre-deliberation straw vote and to their straw vote certainty. Those who were in the low comprehension group were significantly more likely to cast their straw vote for death than were those in the high comprehension group (59 vs. 50%, Chi-square = 3.85; df = 1; $p = .05$; Cramer's $V = .09$). Low comprehension jurors also were significantly more certain of their straw votes in favor of death ($F(1,535) = 4.01, p < .05$; partial eta squared = .01).

Although there was no overall effect of comprehension level on final verdict choice, a linear mixed model hierarchical analysis of the difference in verdict certainty for low comprehension jurors (mean = 5.71) vs. high comprehension jurors (mean = 5.33) nested within juries was significant ($F(1,468) = 5.12$, $p < .05$). Level of instructional comprehension was also significantly related to the way in which the jurors used certain types of penalty phase evidence. A series of linear mixed model hierarchical analyses, using the dichotomous high/low comprehension variable as the independent variable and the eight aggravating and mitigating factors as dependent variables revealed that there was a significant difference in the weight given to two mitigators—the defendant's history of child abuse and the fact that he had pre-existing psychiatric problems—as a function of the jurors' level of instructional comprehension. Specifically, as illustrated in Table 4, high comprehenders weighed these two mitigators more in favor of life than did those jurors with low comprehension scores. Comprehension also affected the way one piece of aggravation—the defendant's prior crime—was used, with high comprehenders weighing it significantly more in favor of death.

Table 4 Effect of level of comprehension on use of penalty phase evidence

Evidence	Low comprehension	High comprehension	*t*-value for fixed effect
Mitigating evidence			
Childhood abuse	3.45	3.12	3.12**
Substance abuse	3.98	3.79	1.64
Psych. problems	3.71	3.37	3.14**
Loved by family	3.64	3.46	1.88^{+}
Aggravating evidence			
Murder facts	5.57	5.74	−1.12
Victim impact	4.80	4.57	1.73^{+}
Prior crime	5.06	5.35	2.19*
No remorse	5.36	5.43	−.57

** $p < .01$; * $p < .05$; $^{+}$ $p < .1$, marginally significant, using linear mixed model hierarchical analysis, with comprehension group nested within jury

Higher number means weighed more heavily toward death; 4 equals no weight given at all

Table 5 Use/misuse of evidence by comprehension level

	Low comprehension			High comprehension		
	Weighed toward life	No weight in decision	Weighed toward death	Weighed toward life	No weight in decision	Weighed toward death
Mitigating evidence						
Childhood abuse**	41% (103)	43% (106)	16% (40)	56% (162)	28% (81)	16% (45)
Substance abuse*	25% (62)	48% (120)	27% (67)	31% (90)	48% (137)	21% (60)
Psych. problems^{+}	33% (81)	51% (128)	16% (40)	42% (122)	46% (133)	11% (32)
Loved by family	32% (80)	56% (131)	12% (31)	40% (115)	52% (149)	8% (24)
Aggravating evidence						
Murder facts	12% (29)	9% (22)	79% (196)	8% (24)	8% (24)	83% (240)
Victim impact	11% (27)	36% (89)	54% (134)	12% (36)	40% (114)	48% (138)
Prior crime*	17% (41)	15% (37)	69% (170)	10% (29)	18% (51)	72% (208)
No remorse*	13% (32)	15% (38)	72% (180)	7% (21)	23% (65)	70% (202)

Low comprehension group $N = 249$

High comprehension group $N = 288$

$^{+}$ $p < .10$; * $p < .05$; ** $p < .01$, using hierarchical linear modeling, with juror nested in jury

Similarly, as Table 5 illustrates, the tendency to properly use specific pieces of evidence varied by level of comprehension. Not surprisingly, jurors who understood the instructions better were more likely to use every piece of evidence properly (i.e., weighing mitigating evidence in favor of life and aggravating evidence in favor of death) than were those jurors whose instructional comprehension was low. Correspondingly, the tendency to *misuse* penalty phase evidence was greater among those whose instructional comprehension was poorer.

The Interplay Between Race, Comprehension, and Verdict

As in our previous study, the jurors' level of instructional comprehension was significantly related to the role that race played in their sentencing decisions. First, comprehension level was related to the racial disparities that occurred in the straw votes, but only in the condition in which a Black defendant was convicted of killing a White victim [Pearson Chi-square $(1, 141) = 6.72$, $p = .01$; Cramer's $V = .22$]. Low comprehenders were significantly more likely to sentence the Black defendant in this condition to death than were the high comprehenders, and the absolute percentage difference was quite large—70 vs. 48%. Instructional comprehension continued to play a role in the race-based death sentencing that occurred in the post-deliberation final verdicts. Specifically, low comprehension jurors were marginally more likely to express more certainty toward a death verdict when they viewed the Black defendant in comparison to those who viewed the White defendant ($F(1,99) = 2.93$, $p < .10$). Overall, 75% of jurors in the low comprehension group voted for death for the Black defendant as compared to 63% who voted for death for the White defendant. Within the high comprehension group, the Black defendant was still sentenced to death at a higher rate than the White defendant (68 vs. 60% respectively), but the margin was smaller.

When only the cross-racial conditions were compared, the influence of instructional comprehension on the final verdicts was even more apparent. Thus, only 56% of jurors with low comprehension voted for a death sentence in the condition where the defendant was White and the victim Black, as compared to fully 75% who voted for death when the defendant was Black and the victim White. For jurors with high instructional comprehension scores, the differential rate of death sentencing was less pronounced (55% for the White defendant/Black victim condition vs. 66% for the Black defendant/White victim condition). Hierarchical linear mixed model analyses of the verdict certainty scores for the two cross-racial conditions indicate that they significantly differ as a function of comprehension ($F(1,232) = 4.42$, $p < .05$).

We also found that instructional comprehension significantly affected the way that jurors used the different kinds of aggravating and mitigating evidence in the Black defendant conditions. Specifically, hierarchical linear mixed model analyses revealed that jurors who judged the Black defendant cases and who had low instructional

comprehension were significantly more likely to weigh three important pieces of mitigation—testimony about the defendant's child abuse, his having had a troubled psychiatric history, and the fact that his family still loved him—in favor of *death* compared to their high comprehension counterparts. In contrast, for those who saw the White defendant cases, only the weighing of evidence about childhood abuse significantly differed as a function of comprehension level.

DISCUSSION

We characterized the pattern of results obtained in our previous study of capital jury decision-making as "complex and disturbing" (Lynch & Haney, 2000, p. 353). Nothing has emerged from the present, more elaborate study to alter our conclusion that the process of death sentencing is fraught with problems, likely to produce unreliable outcomes, and often racially unfair. The primary focus of the current research was to assess the effects of group deliberation on capital jury decision-making and, in particular, to determine whether jury-level group processes lessened or ameliorated the role that race and instructional incomprehension played in death sentencing. Contrary to the findings obtained by some previous researchers—albeit in a non-capital context very different from our own—we did not find that deliberation produced a leniency effect that favored defendants. Deliberation also did not correct the flawed and often confused decision-making process in which the jurors in our simulated penalty trial engaged. Finally, rather than effectively reducing juror bias (in this case, racial bias), as we have noted, deliberation appeared to have exacerbated it.

Deliberation and a Shift Toward Death

In terms of the overall effect of deliberation itself, our jurors shifted some 12 percentage points in their verdict preferences in favor of death verdicts (from 54 to 66%). Thus, in this unique jury decision-making setting where jurors were asked to choose between life and death, our data suggested that the individual jurors were more likely to become *punitive* than lenient when placed in a group deliberation setting. Given the initial slight majority of individual jurors who cast their pre-deliberation votes in favor of the death penalty, at least some of the punitive shift we identified can be attributed to classic group polarization, which occurs most often when the task involves binary judgments (here, choosing between life or death). That is, the tendency for group discussion "to move the average of both individual and group preferences toward whichever pole the pre-discussion individual preference favors" (MacCoun & Kerr, 1988, p. 22).

In addition, we found that the size of the punitive shift varied not only by condition (specifically, as we discuss in more detail below, depending on the racial dynamics of the case), but also as a function of the make-up of the juries themselves. For example, internal analyses of the shift in verdict preferences indicated that White and male jurors started out more in favor of death before deliberations, and became somewhat more in favor of death once deliberations had occurred. However, although women and non-White jurors initially were less in favor of death, and remained so in their final verdicts, they were more likely to shift toward death verdicts in post-deliberation votes. We also found that the proportion of White jurors who were present on a jury elevated its tendency to vote in favor of death overall.

Taken together these findings suggest that the predominance of other pro-death jurors (who tended to be White and male) created considerable momentum in favor of death verdicts that even many initially pro-life jurors (a disproportionate number of whom were women and non-Whites) found difficult to resist. This pattern is not unlike the one identified in studies of actual capital juries that focused on the racial dynamics of the case. Thus, Bowers et al. (2001) identified a "White male dominance effect" at work among the juries whose members they interviewed. Specifically, the presence of a substantial number of White men on the juries increased the likelihood of death verdicts (especially in cases where the defendant was Black and the victim White), leading the researchers to speculate that "[p]erhaps when the jury is dominated by white males, the interpersonal dynamics of decision making is distinctive" (p. 195).

Deliberation and Confusion in Death Penalty Decision-Making

Much previous research has documented the serious and widespread comprehension problems that plague capital penalty instructions (e.g., Diamond, 1993; Eisenberg & Wells, 1993; Wiener et al., 2004), including the instructions that have been in operation in death penalty cases in California over most of the last several decades (Haney & Lynch, 1994, 1997; Lynch & Haney, 2000). Some early research done in a non-capital context suggested that deliberation itself might improve instructional comprehension (Elwork, Alfini, & Sales, 1982), although other research—also done with non-capital instructions—indicated that no net gain in overall comprehension was likely to occur within deliberating juries (Ellsworth, 1989).

The present study was designed to examine whether jury deliberations and the opportunity to discuss and apply capital sentencing instructions as a group enhanced jurors'

comprehension and helped them overcome some or all of the problems that have been so well documented for individual jurors. They did not. In fact, the pattern of post-deliberation comprehension errors we found was strikingly similar to the one identified in our earlier studies of individual-level comprehension. Thus, there was no evidence at all that the opportunity to discuss the instructions and apply them collectively to a set of case facts improved the jurors' comprehension overall. On virtually every dimension that we measured, the jurors' post-deliberation instructional comprehension fell far below what would be expected of rational decision makers whose sentencing verdicts were being properly guided by judicial instructions.

However, although comprehension level did predict the jurors' straw votes (with low comprehenders voting more often for death overall), this pattern was not maintained in the post-deliberation verdicts. This suggests that, although deliberation did not measurably improve the comprehension of the jurors who had the hardest time grasping the meaning of the instructions, it at least brought them closer together in terms of verdicts. On the other hand, this process did not have a salutary effect on each group's use and misuse of the evidence that was presented in the penalty trial. In fact, it may have been that, although the high comprehenders continued to better understand how to properly use the separate pieces of evidence, it did not matter as much; deliberation brought them closer to the low comprehenders in terms of their verdict choice.

Evidence that the jurors were engaged in a confused decision-making process was not restricted exclusively to those jurors whose comprehension was the poorest. As we have noted, there were pronounced tendencies to misuse various case-specific pieces of evidence introduced into the penalty trial the jurors viewed, with an unsettlingly high percentage of them converting aggravation into mitigation and, especially, the reverse.

Of course, it is not surprising that jurors are more likely to appropriately use the evidence that is introduced with the intention of supporting the verdict that they render (i.e., that jurors whose final verdict is death are likely to understand and weigh the aggravating evidence as favoring death, and those whose final verdict is life tend to understand and weigh the mitigating evidence as favoring life). However, the same was not true of the evidence that was introduced to support a verdict that was the opposite of the one that the juror ended up supporting. Although we could not determine the causal direction—that is, whether the differential perception and/or use of the specific category of evidence is what led to the verdict, or the verdict preference influenced the way that the evidence was perceived and/or used—the troubling tendency for jurors misuse penalty phase evidence by actually converting it to its opposite (i.e., regarding mitigating evidence as aggravation and vice versa)—was clear.

Several additional things are worth noting about the way in which key penalty phase evidence was used and misused. One is that, overall, the pattern of misuse was clearly asymmetrical and tended to occur in a manner that would lead to more erroneous or unreliable death rather than life verdicts. That is, we found that the life voters tended to use *all* forms of aggravating evidence more "properly" (i.e., as weighing on the side of death) than the death voters used *any* of the mitigating evidence (i.e., as weighing on the side of life). Moreover, as we have noted, mitigation in general was far more likely to be misused than was aggravation. Thus, a sizable number of jurors (between 14 and 30%, depending on the specific type of mitigation) converted evidence that was intended to be (and statutorily defined as) mitigation into aggravation.

This particular misuse of evidence—the tendency of pro-death jurors to convert mitigation into its opposite—was troubling both because it would speciously inflate the rate of death sentencing and because it is a violation of California death penalty law. That is, although capital jurors in California are free to give mitigation very little or no weight, they are not supposed to use or regard mitigating evidence as aggravation in the weighing process. Indeed, the tendency to misuse the particular kinds of mitigation that we introduced into our simulated penalty trial is especially problematic because precisely this kind of background or "social history" testimony commonly serves as the centerpiece of a capital defendant's case in mitigation (e.g., Haney, 1995). Obviously, when jurors convert background or social history mitigation into aggravation, they are doing more than "merely" ignoring what is commonly intended as the defendant's most compelling argument for life—they are using it as a reason to sentence him to death.

Deliberation and Race-Based Death Sentencing

One of our most troublesome findings was that the deliberation process not only failed to eliminate or reduce the race-based effects that have been identified in many experimental studies of individual juror-level sentencing behavior (e.g., Sweeney & Haney, 1992; Mitchell, Haw, Pfeifer, & Meissner, 2005) but, to the contrary, seemed to activate and amplify racialized decision-making. Thus, the slight but non-significant overall tendency to treat Black defendants more harshly than Whites in pre-deliberation straw votes became larger following deliberation. The post-deliberation shift toward death verdicts that occurred overall was greatest in the two conditions where there was a Black defendant and smallest in the condition in which a White defendant was convicted of killing a Black victim.

Most significantly, our findings that Whites, and particularly White men, demonstrated significant differences based on defendant race in our verdict measures indicates that the "aversive" racism that is activated in this setting is not generic. Rather, it is primarily a problem for White jurors, underscoring the importance of obtaining racially diverse juries in these kinds of cases. Indeed, all of the analyses we performed on non-White jurors in isolation or non-White jurors and women together indicated that the race of defendant did not have a predictive impact on sentence outcomes.

There are a number of potential explanations for the existence of a race-of-defendant effect in death sentencing cases in general (even though some experimental and archival studies have not found it to be present). Long-standing media biases have helped to create and maintain the stereotype of widespread Black criminality (e.g., Entman, 1992). The media also frequently depict Blacks who have been convicted of crime as "irrational, incorrigible, predatory, and dangerous criminals with warped personalities" (Sloop, 1996, p. 116), and generally create a different set of expectations about Black vs. White criminal defendants (e.g., Sunnafrank & Fontes, 1983). In fact, some studies have found that the more stereotypically Black a person's physical appearance is perceived to be, the more likely that person is to be regarded as a criminal (Eberhardt, Goff, Purdie, & Davies, 2004).

Moreover, these stereotypes may account for the public's racialized response to crime in general (e.g., Hurwitz & Peffley, 1997), and may even help to explain race-based differences in the behavior of legal decision makers (e.g., Blair, Judd, & Chapleau, 2004). We know also that when people are faced with an especially "complex judgmental situation"—which death sentencing certainly represents—they rely more heavily on their pre-existing social stereotypes (e.g., Bodenhausen & Lichtenstein, 1987). Indeed, there is now evidence that racial aspects of a defendant's appearance can contribute to him being regarded as "looking deathworthy" (i.e., appearing to deserve a death sentence), at least in cases of cross-racial capital crime (Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006).

However, it is more difficult to explain how and why deliberation per se would increase rather than decrease these tendencies. It is possible that the task at hand—allocating punishment for an undeniably heinous crime—allowed jurors to collectively rationalize whatever amount of punitiveness they were inclined to direct at the defendant, including whatever extra measure was allocated to the Black defendant as a result of pre-existing but unspoken—even non-conscious—racial animus. If the setting itself and the built-in, seemingly impartial justification for maximizing punishment activated aversive racism (Gaertner & Dovidio, 1986) within individual jurors, it would likely be intensified by the presence of others who voiced seemingly race-neutral explanations for their death rather than life verdicts. That is, seeing and hearing others act punitively toward the Black defendant would normalize and legitimize the death verdict, in turn, further liberating or giving permission to other jurors to select it as well.

Sommers and Ellsworth (2003) suggested that race of defendant effects might be more likely to occur in jury settings where the racial issues were not made central or salient part of the case. Thus, they speculated that more "run-of-the-mill" criminal cases were likely to be the ones in which racial bias more often affected the jurors' decision-making process. On the one hand, however, our experimental scenario was hardly a run-of-the-mill crime. On the other hand, the racial identities of the participants—although clearly displayed experimental manipulation—were not particularly highlighted or emphasized in the trial tapes. Moreover, the case facts themselves—those of a robbery murder—were not ones in which racial stereotypes would be particularly engaged.

Devine, Clayton, Dunford, Seying, and Pryce's (2001) comprehensive review of prior jury deliberation research reported "[t]he notable finding… that jury demographic factors interact with defendant characteristics to produce a bias in favor of defendants who are similar to the jury in some salient respect" (p. 673). Aside from gender, of course, race is the most "salient respect" through which jurors can perceive their similarity to or difference from defendants. As we have noted, relatively few of our juries were racially diverse, and fewer still contained Black jurors. This also meant that the typical juror in our study was not afforded an opportunity to engage in the kind of wide-ranging exchange of information during deliberation that Sommers (2006) found was associated with diverse juries and appeared to result in the more lenient treatment of Black defendants.

In fact, these kinds of exchanges of information between diverse participants may be especially infrequent in capital jury deliberations because of the way that the continuing practice of death qualification helps to insure attitudinal homogeneity and minority under-representation (e.g., Fitzgerald & Ellsworth, 1984; Haney, Hurtado, & Vega, 1994). Moreover, minority under-representation on capital cases is often exacerbated by the selective, racially targeted use of peremptory challenges (e.g., Baldus, Woodworth, Zuckerman, Weiner, & Broffitt, 2001). Thus, the racial dimensions to the post-deliberation punitive shift that were found in the present study may closely parallel what happens in actual death penalty cases.

Finally, and not surprisingly, the racial patterns in death sentencing in our study reflected differences in the way jurors used the penalty phase evidence that was presented

in the trial they viewed. The weight given the mitigating evidence by our jurors significantly differed as a function of the racial dimensions of the case. Thus, jurors tended to give mitigation less weight toward life overall for Black defendants, and to give two of four specific pieces of mitigation significantly less weight in their cases when the defendant was Black. Jurors were also more likely to discount two key pieces of mitigation in those cases where the victim was White compared to when the victim was Black, indicating that those in the White victim condition were less receptive to evidence pointing toward a life verdict for the defendant. And, again, there were striking differences in how all of the mitigating evidence and some of the aggravating evidence were evaluated by our White male jurors, as a function of the defendant's race.

We view the racial differences in the use and misuse of mitigating evidence (and the importance of these differentials for death sentencing) as stemming largely from our participants' inability or unwillingness to empathize with the plight of Black defendants, especially in the case where their victim was White. Indeed, they appear to reflect what one of us has termed an "empathic divide"—the relative inability of White jurors to perceive Black capital defendants as "enough like themselves to readily feel any of their pains, to appreciate the true nature of the struggles they have faced, or to genuinely understand how and why their lives have taken very different courses from the jurors' own" (Haney, 2004, p. 1558). A number of previous studies have established the role of empathy—the willingness to take "the perspective of a person in need, imagining how that person is affected by his or her plight (Batson et al., 1997, p. 105)—in the process of assigning punishment. As would be expected, less juror empathy typically leads to harsher punishment recommendations (e.g., Johnson et al., 2002).

This suggests that juries composed largely of White members may legitimize and widen the empathic divide in the course of their deliberations over punishment. Thus, the presence of other more similar-appearing jurors may underscore the defendant's *lack* of similarity, activating racial solidarity rather than cross-racial compassion. In any event, the pattern underscores the special obstacles faced by a Black defendant accused of a capital crime against a White victim, whose case will be judged by a largely or entirely White jury. In this context, the already extremely difficult task faced by capital defense attorneys may be complicated by apparently deep-seated prejudices among at least some jurors that neither judicial instructions nor deliberations appear able to overcome. Attorneys who do undertake this challenge should perhaps be given the legal mandate to employ extraordinary measures to overcome these extra obstacles, as well as the resources with which to do so. Apparently, simply doing the exactly as much for a Black defendant as a White in this situation—equal justice in the form of equal evidence and effort—is not, as we have shown here, enough to produce equal.

CONCLUSION

The present study raises additional questions about the reliability and fairness of the death-sentencing process and, in particular, the behavior of capital juries that are now exclusively entrusted to make the choice between life and death. A large sample of death-qualified, jury-eligible participants deliberated in a reasonably realistic, jury-like setting in which they viewed the videotape of the penalty phase of a capital trial that varied only in terms of the racial characteristics of the defendant and victim. Deliberation shifted the jurors verdict preferences in the direction of the death penalty, did nothing to improve the jurors' overall comprehension of the sentencing instructions intended to guide their exercise of their sentencing discretion, failed to prevent the misuse or improper conversion of mitigating evidence, and appeared to activate and amplify the tendency to treat Black defendants more harshly than their White counterparts, especially among White jurors.

Notwithstanding the considerable strengths that experimental methodology brings to the study of these issues—notably, its unparalleled level of precision, control, and causal clarity—we are well aware that this kind of simulation research also has a number of inherent limitations (e.g., Bornstein, 1999). Although we sought to maximize the ecological validity of our design, we acknowledge that our results are based on a single penalty phase scenario—thus limiting their generalizability. In addition, it is unlikely that our participants experienced the high levels of stress and felt responsibility with which actual jurors grapple in real death penalty cases.

However, properly done simulations clearly add to and supplement data gathered through other methods that have their own different strengths and weaknesses. In fact, there is a high level of consistency between our experimental, laboratory findings and many of the results that have been reported both in archival studies of capital jury verdicts and post-hoc interview studies conducted with capital jurors. Thus, for example, Baldus et al.'s (1998) archival research conducted on a sample of over 500 death-eligible cases in Philadelphia identified several powerful race effects in the capital jury verdicts that were very similar in a number of respects to the ones that we obtained. Specifically, they found that jurors were more likely to find aggravation when the defendant was Black and the victim was not, were less likely to find the existence of mitigating factors in cases where the victim was not Black, and gave mitigating factors less weight when the defendant in the case was Black.

In addition, Bowers and his colleagues in the Capital Jury Project have addressed the question of generalizability in exactly this context. That is, they focused at least some of their path-breaking work on whether and how "patterns of racial bias in mock jury experiments hold for real jurors on real juries in real cases," by conducting an extensive number of post-trial interviews with the people "who actually make the life and death decisions in capital cases" in courtrooms around the country (Bowers et al. 2001, p. 186). Much as we did, they found that racial bias was not reduced or eliminated by the salutary effects of deliberation in cross-racial capital cases. Jurors in those particular cases often failed to engage in a meaningful deliberative process and did not often understand or implement the sentencing instructions they were told to follow. In addition, Bowers et al. (2001) found that the composition of the juries they studied had a "decisive effect on capital sentencing" (p. 195) such that, among other things, juries with higher proportions of White jurors—especially White male jurors—were more likely to reach death verdicts in cases where defendants were Black and victims White.

We conclude by noting that the persistent problems that we and other researchers have documented—through a variety of different research methods—raise troubling questions about in the nature and quality of capital jury decision-making. The legal system's continued unwillingness to both acknowledge these problems and implement effective measures that are designed to eliminate (or at least significantly reduce) their pernicious effects is equally troubling. Together they reflect another significant way in which "our system of capital punishment often falls short of dispensing the fair, consistent, and impartial brand of justice that our Constitution and core political and moral values would seem to require" (Haney, 2005, p. 242).

Acknowledgments We would like to thank our research coordinator, Melissa Bowen, for her invaluable assistance on this project. This research was funded by the National Science Foundation's Law and Social Science Program, Grant #SBR-9996166.

REFERENCES

Abdul-Kabir v. Quarterman, 127 S.Ct. 1654 (2007).

Baldus, D., Woodworth, G., Zuckerman, D., Weiner, A., & Broffitt, B. (1998). Racial discrimination and the death penalty in the post-*Furman* Era: An empirical and legal overview, with recent findings from Philadelphia. *Cornell Law Review, 83*, 1638–1770.

Baldus, D., Woodworth, G., Zuckerman, D., Weiner, A., & Broffitt, B. (2001). The use of peremptory challenges in capital murder trials: A legal and empirical analysis. *University of Pennsylvania Journal of Constitutional Law, 3*, 3–170.

Batson, C., Polycarpou, M., Harmon-Jones, E., Imhoff, H., Michener, E., Bednar, L., et al. (1997). Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of Personality and Social Psychology, 72*, 105–118. doi:10.1037/0022-3514.72.1.105.

Bernard, J. (1979). Interaction between the race of the defendant and that of jurors in determining verdicts. *Law and Psychology Review, 5*, 103–111.

Blair, I., Judd, C., & Chapleau, K. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15*, 674–679. doi:10.1111/j.0956-7976.2004.00739.x.

Bodenhausen, G., & Lichtenstein, M. (1987). Social stereotypes and information processing strategies: The impact of task complexity. *Journal of Personality and Social Psychology, 52*, 871–880. doi:10.1037/0022-3514.52.5.871.

Bornstein, B. (1999). The ecological validity of jury simulations: Is the jury still out? *Law and Human Behavior, 23*, 75–91. doi:10.1023/A:1022326807441.

Bowers, W., & Steiner, B. (1999). Death by default: An empirical demonstration of false and forced choices in capital sentencing. *Texas Law Review, 77*, 605–717.

Bowers, W., Steiner, B., & Sandys, M. (2001). Death sentencing in Black and White: An empirical analysis of the role of jurors' race and jury racial composition. *University of Pennsylvania Journal of Constitutional Law, 3*, 171–274.

Butler, B. (2007). The role of death qualification in jurors' susceptibility to pretrial publicity. *Journal of Applied Social Psychology, 37*, 115–123.

California v. Brown. 479 U.S. 538 (1987).

Carretta, T., & Moreland, R. (1983). The direct and indirect effects of inadmissible evidence. *Journal of Applied Social Psychology, 13*, 291–309. doi:10.1111/j.1559-1816.1983.tb01741.x.

Cowen, C., Thomspon, W., & Ellsworth, P. (1984). The effects of death qualification on jurors' predispositions to convict and on the quality of jury deliberation. *Law and Human Behavior, 8*, 53–79. doi:10.1007/BF01044351.

Devine, D., Clayton, L., Dunford, B., Seying, R., & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy, and Law, 7*, 622–727. doi:10.1037/1076-8971.7.3.622.

Diamond, S. (1993). Instructing on death: Psychologists, juries, and judges. *The American Psychologist, 48*, 423–434. doi:10.1037/0003-066X.48.4.423.

Eberhardt, J., Davies, P., Purdie-Vaughns, V., & Johnson, S. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital sentencing outcomes. *Psychological Science, 17*, 383–386. doi:10.1111/j.1467-9280.2006.01716.x.

Eberhardt, J., Goff, P., Purdie, V., & Davies, P. (2004). Seeing Black: Race, crime, and visual processing. *Journal of Personality and Social Psychology, 87*, 876–893. doi:10.1037/0022-3514.87.6.876.

Eisenberg, T., & Wells, M. (1993). Deadly confusion: Juror instructions in capital cases. *Cornell Law Review, 79*, 1–52.

Ellsworth, P. (1989). Are twelve heads better than one? *Law and Contemporary Problems, 52*, 207–224. doi:10.2307/1191911.

Elwork, A., Alfini, J., & Sales, B. (1982). Towards understandable jury instructions. *Judicature, 65*, 432–443.

Entman, R. (1992). Blacks in the news: Television, modern racism, and cultural change. *The Journalism Quarterly, 69*, 341–361.

Fitzgerald, R., & Ellsworth, P. (1984). Due process vs. crime control: Death qualification and jury attitudes. *Law and Human Behavior, 8*, 31–51. doi:10.1007/BF01044350.

Furman v. Georgia, 408 U.S. 238 (1972).

Gaertner, S., & Dovidio, J. (1986). The aversive form of racism. In J. Dovidio & S. Gaertner (Eds.), *Prejudice, discrimination, and racism: Theory and research* (pp. 61–89). Orlando, FL: Academic Press.

Gregg v. Georgia, 428 U.S. 153 (1976).

Haney, C. (1995). The social context of capital murder: Social histories and the logic of mitigation. *Santa Clara Law Review, 35*, 547–609.

Haney, C. (2004). Condemning the other in death penalty trials: Biographical racism, structural mitigation, and the empathic divide. *De Paul Law Review, 53*, 1557–1589.

Haney, C. (2005). *Death by design: Capital punishment as a social psychological system*. New York: Oxford University Press.

Haney, C., Hurtado, A., & Vega, L. (1994). "Modern" death qualification: New data on its biasing effects. *Law and Human Behavior, 15*, 619–633. doi:10.1007/BF01499328.

Haney, C., & Lynch, M. (1994). Comprehending life and death matters: A preliminary study of California's capital penalty instructions. *Law and Human Behavior, 18*, 411–436. doi:10.1007/BF01499048.

Haney, C., & Lynch, M. (1997). Debating life and death: An analysis of instructional comprehension and penalty phase arguments. *Law and Human Behavior, 20*, 575–595. doi:10.1023/A:1024804629759.

Hurwitz, J., & Peffley, M. (1997). Public perceptions of race and crime: The role of racial stereotypes. *American Journal of Political Science, 41*, 375–401. doi:10.2307/2111769.

Johnson, J., Simmons, C., Jordan, A., MacLean, L., Taddel, J., Thomas, D., et al. (2002). Rodney King and O.J. revisited: The impact of race and defendant empathy induction on judicial decisions. *Journal of Applied Social Psychology, 32*, 1208–1223. doi:10.1111/j.1559-1816.2002.tb01432.x.

Kerr, N., MacCoun, R., & Kramer, G. (1996). Bias in judgment: Comparing individuals and groups. *Psychological Review, 103*, 687–719. doi:10.1037/0033-295X.103.4.687.

Kerwin, J., & Shaffer, D. (1994). Mock jurors versus mock juries: The role of deliberations in reactions to inadmissible testimony. *Personality and Social Psychology Bulletin, 20*, 153–162. doi:10.1177/0146167294202002.

Kramer, G., Kerr, N., & Carroll, J. (1990). Pretrial publicity, judicial remedies, and jury bias. *Law and Human Behavior, 14*, 409–438. doi:10.1007/BF01044220.

Liebman, J. (2007). Slow dancing with death: The Supreme Court and capital punishment, 1963–2006. *Columbia Law Review, 107*, 1–130.

Lockett v. Ohio, 438 U.S. 586 (1978).

London, K., & Nunez, N. (2000). The effect of jury deliberations on jurors' propensity to disregard inadmissible evidence. *The Journal of Applied Psychology, 85*, 932–939. doi:10.1037/0021-9010.85.6.932.

Lynch, H., & Haney, C. (2000). Discrimination and instructional comprehension: Guided discretion, racial bias, and the death penalty. *Law and Human Behavior, 24*, 337–358.

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

McCord, D. (2005). Lightening still strikes: Evidence from the popular press that death sentencing continues to be unconstitutionally arbitrary more than three decades after Furman. *Brooklyn Law Review, 71*, 797–870.

Mitchell, T., Haw, R., Pfeifer, J., & Meissner, C. (2005). Racial bias in mock juror decision-making: A meta-analytic review of defendant mistreatment. *Law and Human Behavior, 29*, 621–637. doi:10.1007/s10979-005-8122-9.

Morgan v. Illinois, 504 U.S. 719 (1992).

Pettigrew, T., & Meertens, R. (1995). Subtle and blatant prejudice in Western Europe. *European Journal of Social Psychology, 25*, 57–75. doi:10.1002/ejsp.2420250106.

Ring v. Arizona, 536 U.S. 584 (2002).

Rompilla v. Beard, 545 U.S. 374 (2005).

Sloop, J. (1996). *The cultural prison: Discourse, prisoners, and punishment*. Tuscaloosa, AL: University of Alabama Press.

Sommers, S. (2006). On racial diversity and group decision making: Identifying multiple effects of racial composition on jury deliberations. *Journal of Personality and Social Psychology, 90*, 597–612. doi:10.1037/0022-3514.90.4.597.

Sommers, S., & Ellsworth, P. (2003). How much do we really know about race and juries: A review of social science theory and research. *Chicago-Kent Law Review, 78*, 997–1031.

Sunnafrank, M., & Fontes, N. (1983). General and crime-related racial stereotypes and influence on juridic decisions. *Cornell Journal of Social Relations, 17*, 1–15.

Sweeney, L., & Haney, C. (1992). The influence of race on sentencing: A meta-analytic review of experimental studies. *Behavioral Sciences & the Law, 10*, 179–195. doi:10.1002/bsl.2370100204.

Thompson, W., Fong, G., & Rosenhan, D. (1981). Inadmissible evidence and juror verdicts. *Journal of Personality and Social Psychology, 58*, 438–449.

Wainwright v. Witt, 469 U.S. 412 (1985).

Wiener, R., Rogers, M., Winter, R., Hurt, L., Hackney, A., Kadela, K., et al. (2004). Guided discretion in capital murder cases: The role of declarative and procedural knowledge. *Psychology, Public Policy, and Law, 10*, 516–576. doi:10.1037/1076-8971.10.4.516.

Wiggins v. Smith, 539 U.S. 510 (2003).

Witherspoon v. Illinois, 391 U.S. 510 (1968).