Jury Decision Making Research: Are Researchers Focusing on the Mouse and Not the Elephant in the Room?

Narina Nunez, University of Wyoming
Jury Decision Making Research: Are Researchers Focusing on the Mouse and Not the Elephant in the Room?

Narina Nuñez, Ph.D.*, Sean M. McCrea, Ph.D.† and Scott E. Culhane, Ph.D.‡

The concerns of jury research have extensively focused on subject selection, yet larger issues loom. We argue that observed differences between students and non-students in mock juror studies are inconsistent at best, and that researchers are ignoring the more important issue of jury deliberation. We contend that the lack of information on deliberating jurors and/or juries is a much greater threat to ecological validity and that some of our basic findings and conclusions in the literature today might be different if we had used juries, not non-deliberating jurors, as the unit of measure. Finally, we come full circle in our review and explore whether the debate about college and community samples might be more relevant to deliberating versus non-deliberating jurors. Copyright © 2011 John Wiley & Sons, Ltd.

INTRODUCTION

The tension between experimental control and ecological validity is present in many applied psychological research endeavors but is probably crucial in jury decision making studies. There has always been a question about whether scientists can adequately study jury decision making in the lab (see, e.g., Bornstein, 1999; Bray & Kerr, 1979; Diamond, 1997), and researchers typically attempt to add features to their studies (e.g., provide videotaped or audio-taped testimony) to enhance ecological validity. Diamond (1997) noted that there were several methodological considerations that could potentially threaten the validity of any study, such as the use of inadequate sampling, simulations of trials that were inferior in quality, a lack of deliberation by jurors in the true jury form, dependent variables that were inappropriate, missing data from the field that corroborated research findings, and the lack of real consequences because of the role-playing nature of decisions. Recently, sample selection has been seen as an area of growing concern.

Reviewers of grants and manuscripts have become increasingly critical of the use of college student populations as mock jurors in jury decision making studies. In our experience, we have had to add increasing justifications in our manuscripts when we use college student samples, and have added community samples to grant proposals after receiving feedback from reviewers encouraging such inclusion. The fact that researchers who have used college samples in their work typically describe their sample as a research

*Correspondence to: Narina Nuñez, Department of Psychology, University of Wyoming, Laramie, WY, U.S.A. E-mail: narina@uwyo.edu
†Department of Psychology, University of Wyoming, WY, U.S.A.
‡Department of Criminal Justice, University of Wyoming, WY, U.S.A.
limitation in their published work (e.g., Golding, Bradshaw, Dunlap, & Hodell, 2007; McCoy, Nuñez, & Dammeyer, 1999) suggests that psycholegal researchers, in general, are being encouraged to move from college to community samples in their work. The concern about sample selection has some face validity because it is typically community members, not college students, who make up the jury panels. Indeed, in most jurisdictions, college students comprise a very small percentage of the jury pool. Further, college students differ from community samples in a number of ways. They are younger, more educated, less likely to be married, less likely to have served on a jury, etc (Nuñez, Dahl, Tang, & Jensen, 2007). Thus, at first blush it appears that criticisms of college student samples are defensible.

We argue in the present paper that, while these criticisms have some face validity, there is often no theoretical reason to suspect such a difference, and only inconsistent empirical support for such claims. Further, we argue that the debate about college versus community samples has obscured a more important weakness of the jury decision research that Diamond (1997) noted: the fact that the vast majority of the published research concerns non-deliberating mock jurors. We contend that the lack of information on deliberating jurors and/or juries is a much greater threat to ecological validity. In addition, we argue that some of our basic findings and conclusions in the literature today might be different if we had included juries, not non-deliberating jurors, as the unit of measure. Finally, we come full circle in our review and explore whether the debate about college and community samples might actually be more relevant to deliberating versus non-deliberating jurors.

**COLLEGE VERSUS COMMUNITY SAMPLES IN NON-DELIBERATING JUROR STUDIES**

In order to examine the importance of utilizing college versus community samples in juror studies, it might be helpful to conceptualize the juror decision making literature into two different types: *process studies* and *content studies* (see Stroebe & Kruglanski, 1989, for a similar distinction). Process research attempts to define the underlying processes and mechanisms that impact juror decisions. For example, several researchers have attempted to examine how jurors reason about a case and make decisions about guilt or innocence (e.g., Bourgeois, Nuñez, Perkins, & Franz, 2002; Pennington & Hastie, 1986). Others have examined how modifying judicial procedures, such as allowing jurors to take notes during trial or altering jury instructions to reduce racial bias, can help people remember information and make better decisions (e.g., Nuñez, Bourgeois, Adams, & Binder, 2004; Shaked-Schroer, Costanzo, & Marcus-Newhall, 2008). Content research attempts to examine how different case contents might result in different decisions. Examples of this type of research include studies that look at how race of the defendant, type of case, age of the victim, etc. impact verdicts (e.g., Bottoms, Golding, Stevenson, Wiley, & Yozwiak, 2007; Nuñez, McCoy, Clark, & Shaw, 1999). The importance of a college or community sample might differ depending on whether one is conducting process or content research.

Process research by its very nature attempts to identify universal human cognitive, social, or social–cognitive processes that underlie decision making. Moreover, because such work is focused on theory testing, there can be less concern about possible sample differences. For example, in a recent paper, Ruva and McEvoy (2008) explored how
pre-trial publicity affected later decisions and memory for information that was presented at trial. College student participants read a packet of news stories about crimes taken from an on-line news source. Some of the stories contained information about a case that, five days later, would be presented in a videotaped trial. The researchers found that, not only did the pre-trial information affect verdicts, but mock jurors also misattributed some of the information they read in the news stories to evidence presented at trial. That is, mock jurors used the pre-trial evidence in their decisions, but then falsely attributed the source of the information to the trial. The findings fit nicely with psychological theory regarding source monitoring (Johnson & Raye, 1981) and research that demonstrates that information presented in one context can be inaccurately attributed to another context (e.g., Durso, Reardon, & Jolly, 1985; Poole & Lindsay, 1995).

In another example of process research, Wiener et al. (1994) examined how counterfactual thinking affected judgments of negligence in a civil suit. After reading about a case that involved a workplace injury, participants were asked to come up with examples of events that could have avoided the workplace injury, prior to rendering their verdict. The researchers found that engaging in counterfactual thinking (i.e., engaging in an exercise that listed events that would have avoided the injury) altered mock juror’s perceptions of the defendant’s behavior, which in turn affected their verdict. Specifically, the counterfactual thinking altered participants’ views about whether the company’s behavior was normal and whether they had provided adequate care for the worker, which then led to fewer pro-defendant verdicts.

The question that arises is whether Ruva and McEvoy (2008) or Wiener et al. (1994) would have observed qualitatively different results had they used community rather than college samples. There is little theoretical reason to suspect that memory and source monitoring errors or counterfactual thinking operates differently in a college or community sample as a result of differences in education, marital status, and so on. Empirical evidence suggests that these processes are even unrelated to general intelligence (Henkel, Johnson, & De Leonardis, 1998). Unless there is a reason to suspect that differences exist, we argue that the college versus community sample debate is of little relevance to research that examines how underlying psychological processes affect decision making. In the case of source monitoring, the only theoretical reason to suspect such differences would be the higher average age of community samples. As past work has shown, accurate source monitoring is significantly worse among older adults (specifically, age 64 and higher; Henkel et al., 1998; Mitchell, Johnson, & Mather, 2003). As a result, one might suspect that source errors would be stronger among community samples. However, even here the differences that might be observed in an older sample are ones of degree rather than of qualitatively different processing of information. Moreover, because a community sample would be relatively heterogeneous, the more interesting question in our view would be to what extent a jury as a whole would be influenced by the incorrect recollections of individual jurors. We return to this issue below. Although we have focused on the examples of source monitoring and counterfactual thinking, we argue that process research should be less susceptible to criticism of sampling for the reason that these processes are likely to be relatively universal (see also Stroebe & Kruglanski, 1989). In sum, jury studies that test psychological theory may not need to be sensitive to sampling issues; however, the same may not be said for content research.

Content research might yield different results in college versus community samples. Moreover, the goal of content research is more frequently centered around predicting
real-world outcomes, and so issues of sampling are of greater concern. For example, because community samples tend to be older, with more life experiences, they might view people and cases differently. If one studied the impact of elderly witnesses on juror decision making it would be reasonable to expect that young college students might react differently than an older community sample. Because there is reason to suspect that some cases might yield different results among college and community samples, it is important to test this by comparing both types of sample on a variety of court cases. This has certainly been done, and by and large data has yielded surprisingly few community–college sample differences (see, e.g., Bornstein, 1999). In one study that examined perceptions of witnesses, researchers varied the characteristics of an elderly witness (Núñez et al., 1999). Participants were presented with one of three elderly stereotypes: frail senior citizen, grandfather, or elder statesman. The authors collected data from community and college samples, reasoning that perceptions of the elderly would likely differ in the two samples. What they found, however, was that both samples returned more guilty verdicts when the witness was an elder statesman than when the witness belonged to one of the other two elderly categories. There were no community–college sample differences on the dependent variables of interest despite the fact that the college sample was considerably younger ($M_{\text{college}} = 21.51$ years; $M_{\text{community}} = 39.15$).

In another study, researchers examined perceptions of juveniles who had committed murder (Núñez et al., 2007). They presented participants with a short case history vignette of a juvenile defendant, varying in age, gender, relationship to the victim, and abuse history. Participants were asked to recommend either adult or juvenile court for the defendant. Reasoning that a younger college sample might differ from an older sample, Núñez et al. (2007) obtained data from three sources: (1) a college student sample ($M_{\text{age}} = 22.58$), (2) a community sample of residents living in the same community as the students ($M_{\text{age}} = 41.90$), and (3) an internet sample of adults living throughout the country ($M_{\text{age}} = 33.45$). Despite expecting differences, the authors found that sample type did not predict any of their dependent variables. Nor were there any interactions between sample type and the independent variables. In fairness, though no sample differences were found, the study did not actually ask mock jurors to make verdict decisions in the case. Instead it tapped attitudes about the proper jurisdiction for juvenile offenders. If the study had asked participants to render verdicts, it is possible that community–college differences might have emerged.

Though we have described only two content studies that have failed to find college–community sample differences, there are many others. For example, Bornstein, Kaplan, and Perry (2007) examined perception of child abuse in a college and community sample. Cutler, Dexter, and Penrod (1989) examined how expert testimony affected decisions and compared a college and a community sample. Roberts and Golding (1991) examined perceptions of criminal responsibility in a case that raised the insanity defense with a college and community sample. In none of these studies were sample differences found. In fact, in a review of 26 studies that compared college and community mock juror samples, most of which were content studies, Bornstein (1999) found that few studies had any sample differences. More importantly, when college–community differences were found, there was no interaction between sample type and other independent variables. For example, Cutler, Dexter, and Penrod (1989) found that their community sample returned more guilty verdicts than did their college sample. However, both samples were influenced by the primary independent variable:
expert testimony on problems with eyewitness memory. Both college and community samples were more skeptical about eyewitness testimony and returned fewer guilty verdicts when they were exposed to expert testimony.

However, there remain reasonable situations in which students and community members could differ. Imagine a case of child sexual abuse. It is reasonable to expect the defense attorney to utilize peremptory challenges on potential jurors who are parents in the venire. At the same time, prosecutors might seek to retain as many as parents as possible. Here, we have differences in the two possible samples (i.e. community members are more likely to be parents) that may translate into verdict or sentencing differences. Individual characteristics, such as age and education, have an impact on decision making (Mills & Bohannon, 1980). The restricted range of both age and education in college samples would likely prevent researchers from identifying these individual difference variables. Indeed, many possible legal situations have not been tested with both samples. Until empirical research comparisons flesh out various scenarios, we will not know for certain whether the two types of samples differ.

Although we acknowledge that there remain legitimate research questions to be addressed in this area, we suggest that other issues of ecological validity may be of graver concern than the issue of sampling. As the title of our paper implies, jury decision making researchers may have paid too much attention to the mouse in the room and ignored the bigger problem in jury research: the elephant in the room. We argue that sample issues are the mouse and that the bigger problem facing the jury decision making literature, the elephant if you will, is the lack of emphasis on deliberating jurors and juries. However, we also suggest that once researchers start conducting deliberation studies the college versus community sample issue may be much more critical.

**DELIBERATING JURORS AND JURIES**

Despite the fact that our judicial system requires group rather than individual decisions, the majority of the jury research has been done on the juror level and then extrapolated to the jury. Certainly one of the main reasons for this practice is the enormous difficulty and time required for group research. Each group made up of multiple participants represents only one data point. A juror research project that requires 60 participants would require 360 for a jury study (using six-person juries). However, the social psychological literature is replete with examples of how group decisions differ from individual decisions. For example, group judgments tend to be more extreme than are individual judgments, an effect labeled group polarization (Teger & Pruitt, 1967; Isenberg, 1986). Polarization can be caused by the consideration of persuasive arguments or information, particularly when the judgment is inherently ambiguous (Isenberg, 1986; Sherif, 1935). For example, Kaplan (1977) provided mock jurors with trial evidence designed to lead them to the conclusion that the defendant was either guilty or innocent. Participants were then given a list of evidence from the case that supposedly had served as the basis of judgment for other jurors. This information was manipulated to be either consistent or inconsistent with the participant’s initial judgment. Ratings of both guilt and recommended punishment became more extreme when the information provided by “other jurors” was supportive of the participant’s initial recommendation, and became less extreme when the other jurors provided information that was contrary to the participant’s initial recommendation.
Polarization can also occur when individuals seek approval from the group (Asch, 1951; Isenberg, 1986). Individuals may abandon their own opinion simply to avoid conflict with others. Such conformity was demonstrated in a series of classic studies conducted by Asch (1951). Individuals were asked to make a series of easy perceptual judgments after hearing the (incorrect) judgments of a number of confederates. Participants reported agreement with the majority on over a third of the trials, despite the fact that the group’s judgment was obviously wrong. In extreme cases, dissent may be met with explicit disapproval or social sanction (Janis, 1982).

In many jury situations, conformity and polarization are likely to occur because of the requirement of a unanimous verdict. Indeed, instructions to reexamine one’s views in order to come to a unanimous verdict increases conformity pressure on minority dissenters (Kassin, Smith, & Tulloch, 1990). It is perhaps not surprising then that past research has typically found a “majority-rules” principle at work among juries, in which verdicts tend to reflect the majority of predeliberation recommendations (e.g., Davis, 1973; Devine, Clayton, Dunford, Seying, & Pryce, 2001; Kalven & Zeisel, 1966). Although one might be tempted to argue that the majority-rules finding indicates that deliberation is uninteresting or unimportant, we hold this view to be misguided. If jurors never changed their vote as a result of deliberation, the majority of cases would end in a hung jury. Devine et al. (2001) reported that first ballots were unanimous in only 35% of the cases examined in their review of the literature. Nonetheless, only 34% of cases with initial disagreements ended in a hung jury. In other words, deliberation played a role in deciding more than 42% of all cases. Moreover, it does not seem to be the case that one can straightforwardly extrapolate jury verdicts from the distribution of juror recommendations. A meta-analysis of jury studies conducted by MacCoun and Kerr (1988) and subsequently replicated by Devine et al. (2001) shows that there is an asymmetry in eventual verdicts in favor of acquittal, such that majorities favoring conviction are overcome in a nontrivial number of cases. A further complicating factor is that, despite any strong recommendations for a unanimous verdict, juries have the possibility of being hung, which is not an option afforded individuals (Davis, 1973).

Thus, an interesting question for future research is to document how influence occurs (i.e., due to either normative social pressure or the sharing of information) and to determine which factors may lead to increased minority influence or stalemate within juries.

In addition to these considerations of social influence within groups, group decisions also differ from individual judgments in terms of process and, as a result, accuracy. For example, group discussions are subject to process loss (Steiner, 1972). That is, important pieces of information may not be shared with the group for various reasons, including failure to communicate effectively (e.g., talking over one another, forgetting what one was going to say, etc.), or social loafing. Increased group size reduces the perceived responsibility any one individual feels for contributing to the discussion, and so vital information may not enter into the discussion (Latané, Williams, & Harkins, 1979, Stasser & Titus, 1985).

Along similar lines, groups may fail to appropriately weight the views of more competent members (Sorkin, Hays, & West, 2001). Applying signal-detection theory to the case of group decision-making, Sorkin et al. (2001) could show mathematically that the performance of a group that fails to deliberate will be considerably lower than ideal, particularly when a unanimous verdict is required. These so-called “Condorcet” groups fail in large part because equal weight is given to both incompetent and competent members. When Sorkin et al. (2001) allowed group interaction in
completing a perceptual task, participants gave weight to member’s judgments in accordance with their actual expertise at the task, and group performance improved. Thus, to the extent that groups are able to adequately share information and identify more competent members, they can come to decisions that are qualitatively better than those of individuals (see also Davis & Harless, 1996). Important for the current discussion, these findings suggest that the judgment of a jury may differ substantially from those of any one juror. Indeed, there are several studies in the psycholegal area that directly demonstrate differences between juror and jury decisions.

In one study, London and Nuñez (2000) examined jurors’ ability to ignore inadmissible evidence in their decision making. At the time the study was done, the conventional thinking was that jurors could not ignore inadmissible but powerful evidence. There were many studies that demonstrated that jurors allowed inadmissible evidence to creep into their judgments (e.g., Kassin & Sommers, 1997). However, most of the research had been done without deliberation. A meta-analysis of studies of inadmissible evidence, examining nearly 50 studies and over 8000 participants, found very few instances of deliberation (Steblay, Hosch, Culhane, & McWethy, 2006). As evidence presented should be the sole consideration for determining guilt or awarding civil judgment, it seems strange that so many different laboratories would ignore this crucial process when determining the impact of inadmissible evidence on decision making.

The question asked by London and Nuñez (2000) was whether group deliberation would enhance or suppress pre-deliberation biases. In two studies, mock jurors were assigned to either an admissible, inadmissible, or control condition. The defendant was charged with taking nude pictures of a neighborhood child, and during trial it was revealed that pictures had been found (Study 1) or that the defendant had been accused previously of committing the same crime (Study 2). Prior to deliberation, jurors in the inadmissible conditions yielded almost as many guilty verdicts as did mock jurors who were exposed to the same information but who were told that they could use the information (i.e., the admissible condition). After deliberations, however, the number of guilty verdicts in the inadmissible condition dropped significantly. In both studies, the decisions of deliberating juries in the inadmissible conditions were not significantly different from those of juries in the control condition who did not receive the damning evidence. Deliberations of the mock jurors eliminated the influence of inadmissible evidence on judgments, in direct contrast to the effects found at the individual level.

In another study, researchers examined whether deliberations led to superior reasoning about a case (McCoy et al., 1999). Participants watched a video of a murder trial and were later asked to reason about the case using the types of question that Kuhn and her colleagues (Kuhn, Weinstock, & Flaton, 1994) have used to study reasoning ability. According to Kuhn et al. (1994), ability to consider alternatives, reflect on evidence, and provide examples of evidence that runs counter to our own judgments are signs of higher level reasoning. Thus participants watched a murder trial and were asked to decide a verdict of first degree or second degree murder, voluntary manslaughter, or not guilty/self-defense. They were asked to determine their verdict and give reasons for choosing the verdict and then to generate the evidence for and against the other verdict options. Jurors were randomly assigned to (1) answer reasoning questions after viewing the trial, (2) answer reasoning questions after deliberating with a group of people, or (3) answer the reasoning questions after ruminating about the details of the trial and taping the ruminations. The researchers found that jurors who answered the questions
after deliberations had significantly higher reasoning skills than either the pre-deliberation or rumination group. The researchers suggested that talking with others, defending one’s views, and hearing other views helped individuals to reason better about the case. The results of the study suggest that we would underestimate case reasoning skills if we only examined non-deliberating jurors.

In a derivation of a study described earlier, Ruva, McEvoy, and Bryant (2007) examined how pre-trial publicity impacted juror and jury decisions. Participants read a series of news articles about crimes (half of whom were exposed to the details of a murder that they would later be asked to judge). Four days after being exposed to the news articles, participants watched a videotaped murder trial. The authors predicted that the impact of pre-trial publicity that they found in their juror study (Ruva & McEvoy, 2008) would be magnified following deliberations. In other words, the authors predicted that prior biases would be magnified in the individuals who deliberated. What they found instead was that, although pre-trial publicity had the predicted effect on individuals prior to deliberations, deliberations led to fewer guilty verdicts in those conditions. In addition, deliberating jurors in the pre-trial publicity condition answered more items correctly on the source monitoring test than non-deliberating jurors. That is, they were less likely to claim that information presented in the pre-trial news articles was actually presented at trial. Similar results were obtained by Kerwin and Shaffer (1994) in a study of the impact of inadmissible evidence. Whereas individual jurors tended to ignore a judge’s instruction not to consider certain evidence, deliberating juries tended to obey this instruction. As a result, although individual jurors and predeliberation juries made equivalent verdict recommendations, postdeliberation juries made more lenient verdict recommendations. Kerwin and Shaffer (1994) also found that the correlation between personal sentiment and verdict was significantly higher for the individual jurors than for the deliberating juries, suggesting that deliberation led individuals to put aside their biased personal feelings and follow the instructions of the judge.

We have reviewed several studies that explicitly compared deliberating and non-deliberating jurors. All found important differences between the deliberating and non-deliberating jurors. Extant data suggests that the quality of the decision can be enhanced through deliberations. After deliberating, jurors are less biased by inadmissible evidence, reason at a higher level about case material, and may be less influenced by pre-trial publicity. Likewise, the review of juries by Devine et al. (2001) found improvements in studies involving deliberating jurors. Specifically, jurors in deliberating conditions had a better understanding of the legal instructions provided. Though we have only reviewed a few studies, all suggest that deliberations can serve to correct initial individual biases. We do not mean to suggest that deliberation always results in superior decisions. As an example of a study that did not find a significant improvement in jurors after deliberation, Wiener and his colleagues examined understanding of judicial instructions at a sentencing phase of a capital trial for pre- and post-deliberation jurors (Wiener et al., 2004). They found that prior to deliberations jurors’ accuracy of legal definitions was better than chance but still quite low. After receiving state approved pattern instructions and having the chance to deliberate with others, comprehension did not significantly increase. Thus, although deliberation may improve final decision making in some cases, it is not a panacea for bad decision making. Importantly, though, the research appears to show that deliberating jurors are not the same as non-deliberating jurors.
THE POSSIBLE IMPORTANCE OF COLLEGE AND COMMUNITY SAMPLES IN DELIBERATING JURIES

In this final section, we come full circle in our arguments and suggest that, once researchers begin to include deliberations as part of their methods, the college–community sample debate may become more important. In fact, it may be that college versus community sampling becomes crucial once deliberating jurors become the norm. Specifically, there is a body of literature in the social and industrial psychological literature that examines heterogeneity of group composition and group decision making. This literature is relevant because samples of juries drawn from a college population are likely to be more homogeneous than samples drawn from the community at large and differences in homogeneity of mock jury groups may lead to different decision outcomes. If so, this would argue strongly for engaging community samples in the jury research.

Two meta-analyses highlight some of the major findings from the homogeneity/heterogeneity literature. Horowitz and Horowitz (2007) examined the literature on team diversity and team outcomes, specifically testing whether bio-demographic diversity (e.g., age, race, gender) or task diversity (e.g., expertise, cognitive ability) impacted group decisions. The researchers identified 35 peer reviewed articles that included 78 different correlations that could be tested. The researchers found that bio-demographic diversity (i.e., variation in demographic traits) was not a significant predictor of quality or quantity of group decisions, but that higher task diversity predicted better decisions. That is, groups that were heterogeneous in terms of expertise and experience typically performed better than less diverse teams. In a somewhat similar meta-analysis, Devine and Philips (2001) examined whether the cognitive ability of individual group members predicted quality of group decision making. They tested whether cognitive ability (mean group member score, highest member score, lowest member score, and standard deviation of scores) predicted team performance. They found that mean group score had the strongest positive relationship with team performance. Low and high member scores also predicted team scores but not as strongly. Interestingly, standard deviation of scores (which would have indicated a more heterogeneous group) was unrelated to team performance.

The burgeoning literature on group homogeneity/heterogeneity recently led van Knippenberg, De Dreu, and Homan (2004) to develop the Categorization Elaboration Model, outlining the circumstances under which heterogeneity is likely to improve or undermine group performance. The model holds that heterogeneity can improve group decision making by bringing a greater range of information and expertise to the task, fostering greater elaboration of relevant information. In a jury setting, heterogeneity could involve a range of relevant experiences, such as prior service as a juror, similarity of experience to the defendant or the victim, or expertise with regard to aspects of the case. On the other hand, diversity can also create intragroup divisions, which undermine these same processes. For example, diversity in gender, age, race, or socioeconomic status may segregate juries in such a way that they refrain from information sharing and fail to elaborate on task-relevant information. Research examining the effects of heterogeneity in the area of jury decision making is less common, but recent findings suggest that there may indeed be important effects of jury diversity on the deliberation process.

In one study, Golding, Bradshaw, Dunlap, and Hodell (2007) investigated how gender composition of a jury impacted decisions. The literature has consistently found
that, in child sexual abuse cases, women mock jurors are consistently more pro-victim (e.g., render more guilty verdicts, believe the victim) than are men (for a review see Bottoms et al., 2007). Golding and his colleagues (2007) tested whether majority men (one woman and five men) and majority women (five women and one man) juries would arrive at different decisions. Specifically, they predicted that majority women juries would render more guilty verdicts than would majority men juries, and that people in majority women juries would be more likely to change a pre-deliberation not guilty vote to a guilty vote after deliberation than people in majority men juries. As expected, they found that women were more pro-defense than were men prior to deliberations. Consistent with their prediction, they also found that, after deliberation, majority women juries returned guilty verdicts more often than majority men juries. Golding and his colleagues (2007) were quick to mention, however, that the women majority juries had an asymmetric pre-deliberation bias towards guilt (i.e., the majority were already leaning towards guilt). As discussed above, if a majority of group members have a pre-deliberation bias towards one outcome, the group decision often reflects that majority (MacCoun & Kerr, 1988). Thus, it may not have been gender of the majority, per se, that determined the number of guilty verdicts in the study. Rather, the distribution of pre-deliberation views may have biased the ensuing deliberation process.

In a similar vein, Sommers (2006) explored racial diversity of the jury and how it impacted decisions and deliberations in the trial of a Black defendant. Though no differences were found in group verdict, compared to all White juries mixed race juries discussed more case facts during deliberations, and this was true for both the Black and White mock jurors. Also, inaccuracies that were brought up during deliberation were less likely to be corrected in the all White juries than in the mixed race juries. Thus, the results suggest that heterogeneity (at least with regard to racial diversity in a trial of a Black defendant) could lead to greater information sharing and superior deliberations.

Finally, Gastil, Burkhalter, and Black (2007) conducted a jury study that more directly addressed the issue of group homogeneity/heterogeneity. A sample of 267 community members, who were to serve on deliberating juries, were recruited to complete a survey both before and after their jury service. Pre-deliberation questionnaires assessed such things as educational attainment, political affiliation and partisanship, political knowledge, and political self-confidence. Post-deliberation questions assessed the quality of the deliberations (e.g., “Jurors thoroughly discussed the relevant facts of the case.” and “All of the jurors listened respectfully to each other during deliberations.”). They also assessed juror satisfaction with the deliberations and final verdict. Though the researchers predicted that diversity (partisan, educational, etc.) would lead to poorer deliberation evaluations, because such diversity would impair the ability of the jury to reach common ground in their deliberations, the results did not support their predictions. The ideological diversity of the jury was not related to deliberation experience and diversity did not prevent careful deliberation.

The body of work described in this section suggests that heterogeneous groups might deliberate and make decisions differently from homogeneous groups. In some instances homogeneity has been found to lead to superior decisions (e.g., leading to less interpersonal conflict; Devine, 1999) and in others heterogeneity has been found to lead to superior decisions (e.g., leading to a greater exchange of information; Sommers, 2006). What is clear, however, is that heterogeneity/homogeneity is relevant to college versus community sampling, regardless of whether one is defining homogeneity by race,

cognitive ability, life experience, etc. In general, we would expect that college samples would be more homogeneous and less heterogeneous than community samples. To the extent that a particular case might favor homogeneity, one might expect that decisions by a group of college students would be superior to those by a group of community residents. To the extent that a particular case might favor heterogeneity, one might expect decisions by a group of community residents to be superior to those by a group of college students. The important point here is that there is reason to suspect that group decisions by community and college samples would differ.

The good news is that research that has shown homogeneity/heterogeneity differences typically finds that group composition effects are small (e.g., Devine & Philips, 2001). However, since most of the research has been done on topics that are relevant to group decisions in the business world (e.g., create a strategic business plan for a fictional regional airline; Devine, 1999) there is a dearth of information about jury decision making. Clearly, more research needs to be done, and research from the management field as well as recent theoretical models such as CEM may provide a guide to researchers interested in these important questions.

CONCLUSION

In this paper, we argue that differences between college and community samples in juror decision research are inconsistent at best. Data supporting community–college differences is not compelling in either the process or content juror research. We argue that a more glaring problem with our body of work is the dearth of research that includes deliberating mock jurors.

We congratulate those researchers who have dedicated their work to the strenuous task of investigating variables at the group level. However, such research comprises a minority portion of legal decision making research. Devine and colleagues (2001) summarized 45 years of jury decision making research and found 206 published articles that included deliberation. Though this is encouraging, it is still an average of only 4.6 articles a year. They recommended that researchers use randomly selected members of the local jury pool as participants in future studies. Yet, their review noted that most studies did not find a difference in the two types of sample.

Once we start employing deliberating jurors in our study, the importance of sampling (college versus community) may indeed become salient. If so, researchers will need to incorporate deliberating community members in their work. We have no doubt that this possibility causes a collective groan among jury researchers. It is difficult enough to recruit community members without also having them deliberate. Though this would make jury research even more difficult to conduct, we may need to do so in order to truly test how process and content variables are likely to impact decision making in real juries.

REFERENCES


