

Law and Human Behavior, Vol. 24, No. 3, 2000

Permitting Jury Discussions During Trial: Impact of the Arizona Reform

Paula L. Hannaford,^{1,4} Valerie P. Hans,² and G. Thomas Munsterman³

A field experiment tested the effect of an Arizona civil jury reform that allows jurors to discuss evidence among themselves during the trial. Judges, jurors, attorneys, and litigants completed questionnaires in trials randomly assigned to either a Trial Discussions condition, in which jurors were permitted to discuss the evidence during trial, or a No Discussions condition, in which jurors were prohibited from discussing the evidence during trial according to traditional admonitions. Judicial agreement with jury verdicts did not differ between conditions. Permitting jurors to discuss the evidence did affect the degree of certainty that jurors reported about their preferences at the start of jury deliberations, the level of conflict on the jury, and the likelihood of reaching unanimity.

INTRODUCTION

“If there was a question right then and there you could talk about it, whereas you didn’t have to wait for nine days or eight days and have to try to jar somebody’s memory. When there was something came up, if you’re in the trial, and we’d talk about it, we all came to a general consensus, this is what it was. You might have wrote it down wrong, you might have heard it wrong, but this is what it was. And then we could go on. We kinda put that as a brick in the wall. Once it was there it was cemented in, we didn’t have to worry about that brick because it wasn’t going to change. Then we’d go on to the next topic and just add the brick.”

Arizona juror interviewed about his experience discussing evidence during trial with other jurors (1997)

The quote above vividly illustrates both the fondest hope and the greatest concern about allowing jurors to discuss the evidence with one another during the

¹National Center for State Courts, Williamsburg, Virginia.

²University of Delaware, Newark, Delaware, and National Center for State Courts, Williamsburg, Virginia.

³Center for Jury Studies, National Center for State Courts, Arlington, Virginia.

⁴Correspondence should be addressed to Paula Hannaford, National Center for State Courts, 300 Newport Avenue, Williamsburg, Virginia 23185 (e-mail phannaford@ncsc.dni.us).

trial. Most jurisdictions forbid jurors to talk together about the case they are hearing until the final deliberations commence, fearing that such conversations will cement the bricks of biased and premature judgment. Yet the juror's description of trial discussions also captures the value of timely and active discussion with one's peers, addressing questions and clarifying evidence while memories are fresh.

In 1995, Arizona adopted comprehensive changes to its jury system. One of the most controversial is Arizona Rule of Civil Procedure 39(f) allowing Arizona civil jurors to discuss the evidence among themselves during the trial (Lakamp, 1998; Myers & Griller, 1997; Shtabsky, 1996). This article reports the results of a field experiment designed to examine and evaluate this legal reform. We first provide theoretical and legal background about jury trial discussions, laying out the most common arguments for and against the reform. After listing the hypothesized effects, we present the field experiment results of the reform's impact.

The Arizona Jury Reforms

Against a backdrop of jury reform initiatives across the country (Munsterman, 1996), the Arizona Supreme Court Committee on More Effective Use of Juries began a comprehensive evaluation and overhaul of its jury system in 1993. In its recommendations, the Committee advocated changing procedures so that the trial was an educational process for jurors and so that jurors had a more active voice in the proceedings. One proposed change was a reform that would explicitly allow civil jurors to discuss the evidence among themselves during trial. The committee's report cited the following benefits: improving jury comprehension; submitting questions and sharing impressions on a timely basis; testing individual jurors' tentative and preliminary judgments against the group's knowledge; and reducing the formation of divisive cliques or forbidden conversations among jurors (Arizona Supreme Court Committee on More Effective Use of Juries, 1994).

When the Arizona Supreme Court approved the reform, Arizona became the only jurisdiction to date that expressly permits civil jurors to discuss the evidence during the trial (Arizona Supreme Court Orders No. R-94-0031 and No. R-92-0004, 1995). In civil cases, judges are to instruct jurors that they may discuss the evidence among themselves if all of them are present in the jury deliberation room. However, jurors are admonished to avoid discussing who should win the case and to refrain from reaching a decision in the case until all of the evidence, instructions, and arguments have been presented (Ariz. R. Civ. P. 39 (f)).

Legal Issues

Although trial judges in some cases have told the jury that discussions among jurors are acceptable, a review of published judicial opinions that have considered this issue shows that most appellate courts criticize the practice (see Hans, Hannaford, & Munsterman, 1999, for a full discussion). Most of these cases, however, have only discussed this issue in the context of juror discussions that took place despite judicial admonitions not to do so—a situation that is distinct from that of Arizona civil jurors (compare *Winebrenner v. United States*, 1945; *People v. Monroe*,

1978; *State v. Washington*, 1980; *Commonwealth v. Kerpan*, 1985, with *Wilson v. State*, 1968; *Meggs v. Fair*, 1980; *United States v. Meester*, 1985). Most of the published opinions are in criminal cases, although the wisdom of jury trial discussions in civil cases has been debated occasionally (*Hunt v. Methodist Hosp.*, 1992).

One of the fullest legal treatments of the potential negative effects of juror discussions may be found in *Winebrenner v. United States*, a 1945 criminal case involving fraud in the procurement of government contracts. According to the *Winebrenner* court, premature judgment is the most serious effect of allowing jurors to discuss the evidence among themselves (*Winebrenner v. United States*, 1945). Other cases follow *Winebrenner* in asserting the dangers of premature judgment, concluding that allowing trial discussions favors the party that goes first (*Commonwealth v. Kerpan*, 1985; *People v. Blondia*, 1976; *People v. Hunter*, 1963; *People v. Monroe*, 1978; *State v. Washington*, 1980).

In contrast, in some cases judges have expressed the view that trial discussions may be beneficial to the jury's decision making. In the most comprehensive defense of juror discussions in a published judicial opinion, Judge Ditter disputed the underlying assumptions on which courts historically had based prohibitions on juror discussions. He argued instead that juror discussions would enhance jurors' attention, comprehension, and ability to recall evidence, thus improving jury performance (*United States v. Wexler*, 1987).

The relevance of the views expressed in these legal opinions to the Arizona situation is debatable. Arizona civil juror discussions are sanctioned by state law, and thus the theme of juror misconduct which runs through many of the published opinions is not pertinent. Judge Dann notes that "[n]either state nor federal case law precludes modifying present practice to allow jurors to discuss the evidence before deliberation" (Dann, 1993, p. 1266). Furthermore, in Arizona a standard instruction describes the conditions under which jurors may discuss the evidence and includes an admonition against reaching premature judgments.

Psychology of Juror and Jury Decision Making

Considering the reform from a theoretical perspective on decision making, what consequences might result from allowing jurors to discuss the evidence during the trial? Some might view jury discussions during trial as inconsistent with the adversary system's traditional passive role for the jury while evidence is presented. In the idealized model of the adversary system, adversary attorneys follow specified rules of evidence in presenting their facts to a neutral, passive decision maker. The passivity of the decision maker is essential in promoting his or her neutrality (Landsman, 1984). Adherence to adversary rules, including the requirement of a neutral and passive factfinder, is thought to promote perceptions of justice (Landsman, 1984; Lind & Tyler, 1988).

In idealized form, the legal model of juror decision making assumes that jurors passively absorb all of the evidence and law presented to them at trial without making any judgments about it until told to do so by the judge (Dann, 1993). In practice the American justice system often departs from the ideals of adversary procedure, including the assumption that jurors refrain from discussing the evidence

until their final deliberations. Indeed, much of contemporary social science research indicates that the model of juror behavior and decision making often described in judicial opinions and legal texts is, for the most part, a legal fiction.

One of the major challenges to the adversary ideal of the passive decision maker comes from psychological research on jury decision making, particularly with respect to social cognition (Fiske and Taylor, 1991). That work indicates that even under traditional adversary rules, jurors take an active approach to their decision making (Diamond & Casper, 1992). The story model of juror decision making (Pennington & Hastie, 1986, 1992) assumes that jurors bring preconceptions and knowledge of the world to their task, that they actively construct narratives or stories from trial evidence, and that they fill in missing details to increase the story's internal consistency and convergence with the jurors' world knowledge.

One of the implications of the story model is that jurors actively make judgments about the evidence and may reject information that is inconsistent with their constructed story (Lord, Ross, & Lepper, 1979). In active information processing, initial impressions shape interpretation of incoming evidence (Fiske & Taylor, 1991; Hastie & Park, 1986). Other research supports the notion that jurors make interim assessments of the evidence and adjust their views as new information is provided (Kassin & Wrightsman, 1988; Weld & Danzig, 1940).

Although mock jury studies provide insight about the factors affecting juror decision making, only a handful focus on group dynamics rather than individual reactions to stimuli (Diamond, 1997; Visher, 1987). A major part of this oversight is the inherent difficulty involved in simulating realistic conditions of jury deliberations. Inadequate time frames for deliberations, failure to provide clear decision rules for verdicts, subjects that are not representative of actual jurors, and the subjects' own knowledge that verdicts have no real effect all contribute to problems of generalizability to actual jury trials.

Fortunately, a few researchers have focused directly on the effect of jury deliberations. Ellsworth (1989), for example, conducted an intensive case study of 18 mock juries in her study of jury deliberation. She concluded that jury deliberation is one of the major strengths of trial by jury and has a salutary impact on its competence. During deliberation, jurors with idiosyncratic views or incorrect recollections of the evidence are confronted with the fact that others on the jury may hold divergent views of the evidence. Jurors must explain, justify, and even defend their views during jury deliberation. Thus deliberation serves as a valuable corrective to idiosyncrasy and error. Ellsworth's findings are consistent with a substantial body of empirical research documenting the advantages of group versus individual decision making, particularly in complex decision tasks (Saks, 1981). Although Ellsworth referred to the traditional final deliberations, one could hypothesize that the opportunity to discuss evidence with other jurors at an earlier stage of the trial might produce similar or even more marked benefits.

Although the process of jury deliberation can be helpful for reducing some forms of bias by individual jurors, it can also amplify the bias of individual jurors under some circumstances. As Kerr, MacCoun, and Kramer (1996) noted in their examination of bias in individuals versus groups, the relative bias of individuals and groups is a function both of the decision-making scheme (unanimity, superma-

majority, or simple majority rules) utilized by the group and of the type of bias (imprecision, commission, omission). They found that the relative strength of decision-relevant information was an important factor in both the direction and the magnitude of individual versus group bias, but that in most studies jury deliberation tended to amplify juror bias with respect to irrelevant or extralegal information (Kerr et al., 1996). Group size, initial individual judgment, and the magnitude of bias among individuals were also found to be important determinants of relative bias in groups. In particular, widely shared biases can be amplified by group discussion (MacCoun, 1995).

The effect of the timing of group discussion is, of course, the key question in this investigation. Lakamp (1998) observes that the timing of the traditional deliberation is not optimally effective in neutralizing the impact of personal preferences and biases, since it occurs after those biases have already developed. She predicts that allowing discussions during trial would be better able to reduce the formation and impact of individual biases before they have been solidified. Studies on modifications of other aspects of jury trial procedure, notably judicial instruction on the law (Kassin & Wrightsman, 1979), demonstrated that timing has a significant impact on juror decision making. Hastie, Penrod, and Pennington (1983) suggest that the timing of activity within group deliberation can have important effects. In their mock jury research, verdict-driven deliberations that were characterized by early polling were shorter, less thorough, and more likely to result in hung juries, whereas evidence-driven deliberations with later polling were more comprehensive and less likely to hang. The present study is the first empirical examination of how the timing of group discussion during the presentation of evidence or afterward affects jury behavior and decision making.

As discussed, there are diverging views of how allowing trial discussions might affect juror decision making. On one hand, it could improve the accuracy and quality of decision making. On the other, it could encourage premature judgment and solidify group biases based on incomplete evidence. Do discussions during trial compromise the neutrality and passivity of the jury? Furthermore, do they undermine the perception of justice?

METHOD

The procedure and materials of the study were designed to address both the potential benefits and the drawbacks of trial discussions. To permit an experimental test of the reform, the Arizona Supreme Court issued an administrative order allowing trial judges in civil trials to depart from normal practice and instruct jurors in some trials that they may not discuss the evidence until final deliberations (Arizona Supreme Court Administrative Order No. 97-1, January 7, 1997).

In the study, civil jury trials were randomly assigned to either an experimental Trial Discussions condition or a control No Discussions condition. Jurors were instructed by the trial judge that they could (or could not) engage in trial discussions depending on the condition to which their case had been randomly assigned. For both conditions, questionnaires were distributed to jurors, judges, attorneys, and

litigants, and a case data form that asked for information about the case was included in the package for judges. Judges in Maricopa, Pima, Mohave, and Yavapai County Superior Courts participated in the study. These four counties account for over 80% of all civil trials conducted in Arizona each year.

The case data form contained factual information about the lawsuit, trial procedures, and case outcome. Judicial questionnaires asked judges to rate the complexity and weight of the evidence, the attorneys' skills, the jury and its verdict, and the trial discussions reform. Juror questionnaires tapped perceptions of the evidence, deliberations, and verdict. Jurors in the Trial Discussions condition also responded to questions about the frequency and nature of trial discussions. Attorneys and litigants were asked about their views of the jury and its verdict, as well as case perceptions and views of trial discussions. Jurors, attorneys, and litigants were also asked demographic questions.

Procedure

A total of 130 packages were distributed to Maricopa County, 100 to Pima County, and 10 each to Mohave and Yavapai Counties, an adequate supply of packages for civil trials for 6 months in those counties. Depending on the random assignment, half of the packages delivered to each court included instructions permitting jurors to discuss the evidence during the trial, while the other half included instructions forbidding the jurors to discuss the evidence during the trial.

In the initial instructions, court staff were asked to distribute materials in every civil jury trial during the study period. They were asked to deliver the page entitled "Instructions to Jurors About Discussing the Evidence During Trial" (which contained the randomly assigned jury instructions) to the trial judge when the jury panel was assembled for voir dire in a civil trial. After the jury was selected, the court staff were asked to give the case data form and the judge's survey to the trial judge. If the case settled or a mistrial was declared before the jury deliberated, the judge was asked to complete the case data form and return it with the uncompleted surveys to the court administrator.

After the jury retired to deliberate, the court staff were asked to distribute the attorney and party questionnaires to lead counsel for each side in the case, requesting that attorneys distribute the questionnaires to the first-named plaintiff or defendant in the lawsuit. After the jury reached a verdict or declared itself hung, the court staff were asked to distribute the juror surveys. The trial judge in the case was given the option of allowing jurors to fill out the questionnaires while waiting for the attorneys and parties to return to the courthouse, or waiting until after the formal verdict was announced in court. Court staff collected all surveys from the judge, jurors, attorneys, and litigants, placed them in the envelope, and gave the envelopes to the court administrator, who then forwarded them to the National Center for State Courts for data entry and analysis.

Data collection began on June 15, 1997. By January 30, 1998, 214 packages had been distributed to a trial judge presiding over a civil jury trial. Of these, all of the questionnaires from 12 packages were not returned by the trial judges, but for 202 cases at least some questionnaires were returned, for an overall response

rate of 94%. From the case data forms and other court sources, we counted the number of settlements, mistrials, and directed verdicts. Subtracting the cases in which juries had not deliberated (e.g., settlements, mistrials, and directed verdicts), there were a total of 192 valid cases in the sample. The total number of 214 cases was used to calculate the case form response rate, since judges were asked to fill it out whether or not the jury deliberated. The 192 valid cases were used to calculate the response rate for the other forms, since judges, jurors, attorneys, and litigants were only asked to complete the forms if the jury deliberated. Response rates for each category of survey respondent are shown in Table 1.

Random Assignment

Trial judges in the experimental Trial Discussions condition were given a sheet headed in bold, "For this trial, please instruct the jurors that they may discuss the evidence prior to deliberations according to ARCP Rule 39(f)." They were invited to use a suggested instruction that we developed on the basis of the pretest, the instruction from the Court Comment for ARCP Rule 39(f), or an instruction of their own choosing. Trial judges in the control No Discussions condition were given a sheet headed in bold, "For this trial, please instruct the jurors that they may NOT discuss the evidence prior to deliberations." They were invited to use our suggested instruction or one of their own choosing.

At least one survey was returned in 172 cases. We confirmed that judges followed the random assignment in 161 cases (94%). In 5 cases (3% of the total), there was evidence that the judge had not followed the random assignment (3 in one condition and 2 in the other). In the remaining 6 cases, it was unclear whether the judge had followed the requested random assignment. These 11 cases were deleted from all analyses in which the impact of Rule 39(f) was assessed.

Data Analysis Strategy

The fact that this was a field experiment using actual jury trials limited the possible methodological approaches. There was no ability to manipulate evidentiary

Table 1. Response Rates for Survey Instruments

Survey type	Cases	Returned	Response rate (%)
Case data form	214	189	88
Judge's questionnaire	192	167	87
Jury questionnaire ^a	192	170	87
	(1,597 jurors)	(1,388 jurors)	(89)
Attorney questionnaire	192	126 Def. 95 Pt.	66 Def. 50 Pt.
Litigant questionnaire	192	71 Def. 50 Pt.	37 Def. 26 Pt.

^aThe number of total possible jurors was estimated as the number of valid cases (192) times the average number of deliberating jurors (8.32). Arizona uses 8-person civil juries, but alternate jurors may participate in the final deliberations with the consent of the parties (Munsterman et al., 1997).

factors in the trials or even the range of case types included in the sample. The degree of variation among the sample cases made it highly impractical to devise an objective test of juror comprehension (e.g., accuracy of juror recall for evidence or understanding of legal instructions) such as those typically employed in mock jury studies. Moreover, the trials could not be interrupted to collect interim responses from the study participants at different stages of the trial.

Because of these field constraints, the questionnaires were designed to solicit retrospective information from each of the trial participants, despite the risk that this type of information can introduce errors and biases to the data. Of particular significance to the present study is the question of how accurately jurors can assess their own decision-making processes (Nisbett & Ross, 1980). Self-report data provide insight about jurors' perceptions of their performance, but not necessarily about their actual performance.

Survey respondents might also recall events occurring more recently than they actually did, a phenomenon called telescoping. Several measures in the study protocols were included to minimize error, including using individual and anonymous survey instruments, keeping the reference period short (e.g., the duration of the trial), and using trial benchmarks (e.g., opening statements, plaintiff and defendant evidence, closing arguments) as memory anchors (Pearson, Ross, & Dawes, 1992). Even so, juror respondents may have adjusted their responses to show the greatest consistency with the outcome of the trial or to provide responses that they perceived as socially desirable. The great degree of variability in juror responses suggests that this effect may have been negligible.

A number of different approaches were considered to examine the effects of trial discussions. Comparing the experimental and control conditions furnishes information about the overall impact of instituting trial discussions, and the Results section presents these key contrasts of experimental and control conditions. However, for some analyses we compared the cases in which jurors either did or did not participate in the approved discussions during trial. Comparing juries that did and did not discuss can be useful because it may provide insight into how the experience of engaging in trial discussions affects the process and outcome of jury decision making.

Finally, we reviewed methods of describing and analyzing the collective accounts of juror reports and experiences. We concluded that it was most appropriate to calculate and analyze group-level jury data for questionnaire responses. Many of the hypothesized effects of trial discussions are group-level predictions about how the jurors as a collective decision-making body might be influenced. It is also the case that jurors within juries are not independent and to use their individual data would violate assumptions of common statistical analyses. Unless otherwise indicated, the paper uses summary measures (e.g., means, percentages of jurors) with the jury as the unit of analysis.

Predictions about the impact of permitting jurors to discuss the evidence during trial were derived from legal opinions, the Arizona Supreme Court Committee on More Effective Use of Juries' own justifications, and psychological research on juror and jury decision making. The hypothesized effects pertain to the following domains: individual juror decision making, interactions with other jurors, and the deliberation process and jury decision making.

RESULTS

Comparability of the Experimental and Control Conditions

To determine whether random assignment had resulted in two comparable groups of cases, the cases were contrasted on a number of dimensions including study site, the type of claim, the number of witnesses and exhibits for each side, the trial length, the judge’s overall rating of the degree to which the evidence favored the plaintiff or the defendant, and judicial and juror ratings of the complexity of the case. One apparent difference was in perceptions of the complexity of the case. Judges rated the No Discussions cases as more complex than the Trial Discussions cases, $M = 3.45$ for No Discussions, $M = 2.84$ for Trial Discussions; $t(148) = 2.26, p = .025$. Jurors’ ratings followed a similar, though not statistically significant ($p = .20$) pattern. We examined measures of evidence and trial length, and although they pointed in the direction of somewhat longer trials (approximately 2 hr longer on average) with more defense evidence in the No Discussions condition, the measures were highly variable and none differed significantly between the two conditions. We created a Case Complexity Scale (alpha = .78) that included the judge’s rating of complexity, the jurors’ average rating of complexity, and the natural log of the total number of hours of the trial. Trial complexity as measured by the Case Complexity Scale differed significantly by condition, with No Discussions cases more complex than the Trial Discussions cases, $t(144) = 2.37, p = .019$; see Table 2 for means. Thus, the No Discussions cases were by chance somewhat more complex. Key analyses took into account the difference in case complexity, for example, by including it as a factor in regression analyses.

There were also significant differences in the case characteristics by site, particularly between Maricopa County and Pima County, which together comprise 96%

Table 2. Site Characteristics

	Non-Pima counties		Pima County	
	No discussions juries ($n = 47$)	Trial discussions juries ($n = 47$)	No discussions juries ($n = 26$)	Trial discussions juries ($n = 35$)
Judicial ratings of evidence				
Favored plaintiff	15 (32%)	11 (23%)	6 (23%)	17 (49%)
Evenly balanced	19 (40%)	17 (36%)	9 (35%)	9 (26%)
Favored defendant	13 (28%)	19 (40%)	11 (42%)	9 (26%)
Case characteristics				
Case complexity M	10.1	9.1	9.2	7.9
Trial length (hr)	21.7	18.8	14.0	13.5
Caseload composition				
Auto tort	13 (28%)	19 (40%)	17 (63%)	21 (66%)
Premises liability	6 (13%)	6 (13%)	4 (15%)	4 (13%)
Medical malpractice	2 (4%)	3 (6%)	2 (7%)	1 (3%)
Products liability	2 (4%)	4 (9%)	2 (7%)	1 (3%)
Other tort	14 (30%)	6 (13%)	0 (0%)	5 (16%)
Contract	10 (21%)	8 (17%)	2 (7%)	0 (0%)

of the trials in the study (see Table 2). The difference between the sites and the composition of their respective caseloads became critically important with respect to a number of dimensions of our analysis. Case complexity, for example, was significantly lower in Pima County than in the non-Pima counties, $t(137) = 2.39$, $p = .018$. A common belief of both proponents and critics of permitting juror discussions during trial is that the reform's advantages and disadvantages are most likely to be felt in longer, more complex trials. The diverse caseload composition included in these data may mask the effects of the reform—good or bad—in specific types of cases.

Another important issue related to the comparability of the cases occurred in the distribution of judicial assessments of the weight of the evidence in the sites, with significantly more Pima County judges reporting that evidence favored the plaintiff (38%) than judges in non-Pima sites (28%).

A complicating factor was the fact that judicial assessments of the evidence did differ by experimental condition in Pima County, possibly due to chance or to noncompliance with the random assignment protocols in that site. We examined the frequency of settlement, mistrials, nonresponse, and apparent noncompliance with random assignment to assess the extent to which these factors could have affected the balance of cases across conditions in Pima County. We found that 5 cases were omitted from the No Discussions condition in Pima County, whereas only 1 case was omitted from the Trial Discussions condition because the judge did not comply, or we were unable to determine whether the trial judge complied, with the experimental condition. Other factors affecting the difference in random assignment included 4 No Discussions cases being omitted due to mistrials (1) and nonresponse by the trial judges (3) compared to 2 Trial Discussions cases (1 settlement and 1 nonresponse).

As Table 2 illustrates, a much higher proportion of cases in which the evidence favored the plaintiff was by chance assigned to the Trial Discussions condition in Pima County than in the non-Pima sites, $\chi^2(1, N = 49) = 4.98$, $p = .03$. Not surprisingly, this might have a profound effect in the analyses of the juries' verdicts. For other dimensions of the analyses, however, the site differences did not affect our conclusions about the effects of the reform. Therefore, unless indicated otherwise, the statistics reported in this article reflect the combined data from all of the sites.

Of the 84 cases in which jurors were permitted to discuss the evidence, jurors reported that they had no formal discussions in 26 cases (Hans et al., 1999). Whether or not jurors engage in discussions is likely to be influenced by a range of case and other factors and those must be assessed to determine the ways in which these other factors could independently affect the dependent measures of interest. A series of statistical analyses was undertaken to compare simultaneously the control group ($N = 75$ juries), the juries in the experimental group that chose not to discuss the evidence ($N = 26$ juries), and the juries in the experimental group that engaged in discussions ($N = 58$ juries). Post hoc tests were then performed to determine whether case dimensions were significantly different among the three groups. Case complexity again differed among the groups, $F(2, 150) = 5.06$, $p = .008$. The overall Case Complexity Scale measure showed that the cases of juries that could discuss,

but did not were the least complex ($M = 7.54$), the cases of juries that discussed the evidence during trial were in between ($M = 9.06$), and the control cases were the most complex ($M = 9.77$).

Except for case complexity, we found no other significant differences with respect to case characteristics. We did note, however, that the average trial length for the juries that were permitted to discuss the evidence but did not was much shorter compared to the other juries. This was certainly true in Pima County, where 40% of the jurors assigned to Trial Discussion cases did not engage in formal discussions compared to 24% of non-Pima jurors assigned to Trial Discussion cases. We believe that most of the juries who chose not to discuss the evidence may have served in relatively short cases that did not provide substantial opportunities for jurors to talk, although the wide variance in trial lengths obscured any measurable differences between these subgroups (Hans et al., 1999).

A final difference among the cases was that the judge's assessment of the effectiveness of the plaintiff's attorney differed significantly across the three sets of cases, $F(2, 155) = 3.54$, $p = .03$, with the most negative judgments coming in the condition in which jurors were permitted to discuss the case, but did not do so ($M = 4.15$ in those cases, compared to 4.8 and 5.0 in the other two sets of cases). These differences should be kept in mind in evaluating the impact of trial discussions.

Test of Prejudgment

Perhaps the most ubiquitous fear expressed is that trial discussions will cause jurors to reach premature decisions about who should win the case. The danger of prejudgment is mentioned in many legal opinions as a potential hazard of allowing jurors to discuss the evidence among themselves. Psychological research also points to the potential hazards of active information processing. If the critics are right, jurors who engage in trial discussions should report that they started leaning to one side or the other and made up their minds at earlier points than jurors who are not permitted to engage in trial discussions. Alternatively, jurors may be less likely to form or retain early opinions if they are exposed to differing views.

To test these hypotheses, we provided a list of different stages of the trial on the juror questionnaire and asked jurors to indicate in what stage of the trial they had started leaning toward one side or another and at what stage they made up their minds about who should win the case. Possible responses for both conditions included: the plaintiff's opening statement, the defendant's opening statement, the plaintiff's evidence, the defendant's evidence, the plaintiff's closing argument, the defendant's argument, the judge's instructions, discussions with jurors during trial, and final deliberations. Using the percentage of jurors who responded affirmatively to each trial segment as dependent variables in a MANOVA analysis, we found that the reports of when jurors began leaning toward one side, $F(8, 161) = .42$, ns, and when jurors made up their minds about who should win the case, $F(8, 161) = 1.21$, ns, show no significant differences as a function of trial discussions. Adding judicial assessment of the evidence, case complexity, and site as covariates in the MANOVA analyses similarly failed to find a significant effect from trial discussions (all F 's < 1 , ns).

Primacy and Recency Effects

Another way of assessing whether trial discussions encourage premature judgments is to look at whether or not they influence the jurors' perceived importance of the first and last witnesses. The first witness is likely to be the subject of trial discussions and might be more noticed and influential if he or she is the subject of such discussions. This might enhance the typical primacy effect, and increase ratings of his or her influence. In addition, we might expect jurors to pay more attention to the first phase of the trial. The last witness, and the last half of the trial, should be less important if one has already prejudged the case and made up one's mind.

Table 3 compares a number of variables related to primacy and recency by experimental condition. There was no difference between the Trial Discussions and No Discussions juries in their ratings of the importance of the first and last witnesses or their attention during the first half of the trial, but there was evidence of a more pronounced recency effect for the No Discussions juries. These groups rated their memory for the second half of the trial as slightly but significantly higher than the experimental groups. Judicial, attorney and litigant assessments of the juries' attention to first and second parts of the trial did not differ significantly by experimental condition.

Juror Preferences at the Beginning of Deliberations

There were some clear differences in jurors' reports of their views at the beginning of the deliberation. Table 3 shows that jurors who had the opportunity to discuss the case during the trial were less likely to say that they were unsure who they favored at the start of deliberations. There were no differences in the extent to which juries in the two groups favored the defendant at the beginning of the deliberation. Instead, jurors who had the opportunity to discuss the case during trial were significantly more likely to report that they favored the plaintiff at the start of deliberations. Thus it seems that a proportion of jurors in the Trial Discussions condition who otherwise might have been unsure may have shifted to favoring the plaintiff.

Did the discussions themselves produce this orientation, or at least a willingness to express a view at the start of deliberations, or was it an artifact of the set of cases or some other set of variables? Introducing controls into a linear regression, $F(5, 136) = 6.00, p < .001, R \text{ square} = .18$ predicting juror favoritism to the plaintiff at the start of deliberations, including the Case Complexity Scale ($\beta = .18, p = .03$), the judge's rating of the closeness of evidence ($\beta = -.316, p < .001$), site location (Pima/non-Pima), ($\beta = .05, ns$), and the judge's perception of the relative abilities of the attorneys ($\beta = -.01, ns$), does not eliminate the impact of trial discussions ($\beta = .22, p = .006$).

However, examining the three types of discussion groups suggests that whatever effect exists may not be due exclusively to the fact of jurors discussing the evidence. The effect disappears completely when Trial Discussions is replaced as a predictor with a variable indicating whether jurors actually engaged in formal discussions ($\beta = .05, ns$). Table 3 shows that the highest percentage of juries favoring the

Table 3. Comparisons of Trial Discussions and No Discussions Juries

Dependent variables	Trial discussions juries (did/did not discuss)	No discussions juries	<i>df</i>	<i>F</i>
Case characteristics				
Case complexity	8.6 (9.1/7.5)	9.8	1, 149	5.68*
Trial length (hr)	16.7 (18.5/12.4)	18.9	1, 153	.89
Primacy/recency effects				
Juror rating: first witness	5.0 (4.9/5.3)	5.1	1, 157	.05
Juror rating: last witness	5.0 (4.9/5.3)	4.8	1, 157	1.26
Juror: attention, first half of trial	5.8 (5.8/5.7)	5.8	1, 157	.02
Juror: attention, second half of trial	6.1 (6.1/6.0)	6.2	1, 157	4.67*
Judge: jury attention, first half	5.8 (5.8/5.8)	6.0	1, 158	3.10
Judge: jury attention second half	5.5 (5.5/5.6)	5.6	1, 157	.36
Attorney: Jury attention first half	5.6 (5.6/5.7)	5.7	1, 191	.12
Attorney: Jury attention second half	5.3 (5.4/5.1)	5.5	1, 191	.94
Litigant: Jury attention first half	5.8 (6.0/5.4)	5.8	1, 100	.02
Litigant: Jury attention second half	5.3 (5.3/5.4)	5.8	1, 100	2.02
Jury views, start of deliberation				
Favor plaintiff (% of jury)	30 (27/38)	20	1, 157	5.46*
Favor defendant (% of jury)	38 (39/36)	37	1, 157	.07
Unsure (% of jury)	32 (35/26)	43	1, 157	10.28**
Jury votes and case outcomes				
P votes, first ballot (% of jury)	46 (45/48)	37	1, 147	2.59
P votes, last ballot (% of jury)	49 (48/50)	42	1, 157	1.28
Plaintiff verdicts (%)	67 (68/63)	55	1, 158	2.59
Percent unanimous verdicts	32 (29/38)	49	1, 159	4.88*
Judge/jury agreement	82 (75/92)	86	1, 152	.54
Compensatory awards <i>M</i> (\$)	204,156 (264,401/64,141)	210,159	1, 91	.00
Log of awards <i>M</i>	10.2 (10.5/9.6)	10.1	1, 91	.013
Jury comprehension				
Juror: Evidence comprehension	5.4 (5.4/5.4)	5.3	1, 157	.56
Juror: Experts comprehension	5.6 (5.6/5.8)	5.5	1, 156	2.02
Juror: Law comprehension	6.1 (6.0/6.3)	6.2	1, 156	.39
Judge: Jury comprehension	5.9 (5.7/6.2)	6.1	1, 151	1.45
Attorney: Jury comprehension	5.0 (5.0/4.9)	5.4	1, 183	2.11
Litigant: Jury comprehension	4.8 (5.0/4.2)	4.8	1, 91	.01
Jury interaction				
Informal discussions with jurors (% of juries)	31 (35/21)	14.1	157	18.25**
Informal discussions with friends or family (% of juries)	11 (11/11)	14	1, 157	2.53
Level of conflict in deliberation	2.0 (2.0/1.9)	1.8	1, 157	2.02

Note. Means on a scale of 1–7 unless otherwise noted. For the Trial Discussions juries, figures in parentheses are the mean values for juries that reported engaging in formal trial discussions and juries reporting that they did not in formal trial discussions, respectively.

* $p < .05$; ** $p < .01$.

plaintiff at the beginning of the deliberation comes in the groups who are allowed to discuss the evidence, *but do not*. The Trial Discussions juries who actually engaged in discussions are intermediate, while the No Discussions group is lowest. If the discussions themselves produced a plaintiff tilt, one would have expected the highest percentage among the juries who engaged in discussions, but that does not occur.

The same pattern occurs with jurors who are unsure who they favor at the

beginning of deliberations. Here the highest group is the No Discussions condition juries, the lowest is the Trial Discussions juries who did not have any discussions about evidence during the trial, and again the juries who engaged in trial discussions are intermediate. Although the overall *F* tests for favoring the plaintiff and unsure were statistically significant, the post hoc tests showed that the No Discussions condition juries were statistically different only from the Trial Discussions juries who chose not to discuss the case. The Trial Discussions juries who actually engaged in discussions of evidence during the trial were no different from the other two groups.

Jurors' reports of their votes on the first and final ballots, and the final case outcomes, follow the same pattern of somewhat higher plaintiff votes and verdicts in the Trial Discussions condition in analyses of variance (see Table 3). There is a strikingly different pattern among the study sites. In the non-Pima sites, the proportion of juries deciding for the plaintiff is virtually identical for No Discussions and Trial Discussions cases, 59% versus 60%, respectively; $\chi^2(1, N = 99) < 1$, ns. However, in Pima county cases, there is a difference in plaintiff verdicts, with more plaintiff verdicts in the Trial Discussions condition, 46% for No Discussions and 77% for Trial Discussions cases; $\chi^2(1, N = 61) = 6.21, p = .01$. It will be recalled that judicial views of the evidence also followed this pattern. We analyze the jury verdicts further below.

The compensatory awards were highly variable in both conditions, and there were no significant differences between conditions (see Table 3). A second analysis using a natural logarithmic transformation of the compensatory damage awards also produced no significant differences. There were too few punitive damage awards to plaintiffs ($N = 6$), and too few defense judgments, either compensatory ($N = 7$) or punitive ($N = 2$), to permit separate statistical analysis; adding the plaintiff's punitive award to the compensatory award did not alter the overall findings.

Tests of Improvement in Understanding of Evidence and Law

One of the major justifications for allowing trial discussions is to promote jurors' understanding of evidence, particularly complex evidence that might be too difficult for individual jurors to understand without assistance. Jurors may also increase their understanding of judicial instructions, although that may depend on when the substantive instructions are given. If the advocates of trial discussions are correct, jurors who engage in trial discussions should report that they have a better understanding of the evidence (and, possibly, the law) than jurors who do not engage in trial discussions. Furthermore, the judge, attorneys, and litigants in trial discussions cases should be more likely to agree that jurors appeared to understand the evidence presented at the trial.

Jurors in our study reported that trial discussions were helpful (see Hans et al., 1999, for a full presentation). Approximately 8 of 10 jurors who had engaged in trial discussions agreed that trial discussions improve the jury's understanding of the trial evidence. Jurors who engaged in discussions during the trial described them as thorough, accurate, and helpful in resolving confusion. Furthermore, the

majority of jurors who were not permitted to discuss the evidence during the trial said that they would have liked to do so (Hans et al., 1999).

When asked how difficult it was to understand the evidence, the expert evidence, and the judicial instructions on the law in the cases before them, both the Trial Discussions and No Discussions juries were quite positive about their understanding of the evidence and the law. They stated that they had a relatively good understanding of the evidence, the expert evidence, and the judicial instructions (see Table 3). Because case complexity differed between conditions, regression analyses of the impact of trial discussions on juror comprehension included the Case Complexity Scale as a control. We undertook three separate linear regressions in which jurors' reports of understanding the evidence, $F(3, 147) = 8.97, p < .0001, R^2 = .155$, expert testimony, $F(3, 146) = 4.37, p = .006, R^2 = .082$, and judicial instructions, $F(3, 147) = 4.45, p = .005, R^2 = .065$, were the respective dependent variables. The independent variables were the experimental condition, Case Complexity, and site (Pima/non-Pima). Case Complexity was the only significant factor (beta = $-.36, p < .0001$, for understanding of evidence; beta = $-.25, p = .003$, for understanding of expert evidence; beta = $-.26, p = .002$ for understanding of judicial instructions) for all three regressions. As Case Complexity increased, jurors reported more trouble understanding the evidence and instructions. Experimental condition and site variables were not statistically significant (all betas $< .113, ns$).

Judge, Attorney, and Litigant Assessments of Juror Comprehension

Judges also rated the jury's likelihood of understanding the major evidentiary and legal issues in the case about the same across conditions (see Table 3). In general, judges thought juries understood the key issues in virtually all the trials. Judges rated juror comprehension below the midpoint in just 6 of the 153 cases.

Neither the attorneys nor the litigants reported different assessments of jury comprehension as a function of trial discussions (attorney and litigant samples *ts* all *ns*, see Table 3). There was, however, a perhaps understandable link between attorney and litigant assessments of juror comprehension and the verdict. In cases in which the jury returned a favorable verdict, attorneys rated comprehension more highly, compared to cases with unfavorable verdicts, $M = 5.53$ versus $M = 4.78$; $t(183) = 3.01, p = .003$. The verdict figured even more prominently in litigant ratings of juror comprehension, $M = 5.56$ versus $M = 3.77$; $t(91) = 4.08, p < .0001$.

Judicial Assessments of the Jury Verdict

Another angle on the accuracy of the jury's verdict is a comparison of judge and jury views of the case. Ever since Kalven and Zeisel's (1966) landmark study of judge-jury disagreement, scholars have been interested in determining the extent to which judges agree with jury verdicts in the cases over which they preside. In our study, we wanted to discover whether judicial agreement changes when jurors have the opportunity to discuss the evidence during the trial. If the reform improves the jury's ability to evaluate evidence in line with judicially sanctioned legal stan-

dards, then one would predict that judge–jury agreement would be higher, and judges would express less surprise and more satisfaction with the verdict (see Heuer & Penrod, 1989, p. 425, for a similar prediction regarding other jury reforms). This is one interpretation; the exact meaning of judge–jury agreement has been the subject of debate.

Judges were asked to rate, on a 7-point scale, the degree to which the evidence favored the plaintiff (1, 2, 3), was evenly balanced (4), or favored the defendant (5, 6, 7). They were also asked to indicate how surprised and how satisfied they were with the jury's decision in the case. Judge–jury agreement was defined as any instance in which a judge rated the evidence as evenly balanced or favoring the verdict that the jury reached in the case. A judicial rating that the evidence favored the losing side constituted disagreement. This was a conservative measure of disagreement, since in cases in which the evidence was evenly balanced judges might well reach a verdict different from the jury's. In addition, it is possible that some judges may have used the midpoint of the scale to avoid commenting on the weight of the evidence.

Overall, judicial ratings of the plaintiff versus defense slant of the evidence were in line with the outcomes reached by the jury, with ratings that the evidence favored one side higher in the cases in which that side had prevailed. Fully 35% of the judges rated the evidence as evenly balanced. The No Discussions and Trial Discussions conditions produced similar ratings by the judges. Likewise, surprise and satisfaction with the verdict were no different for the two conditions (for surprise: No Discussions $M = 2.97$, Trial Discussions $M = 2.85$; for satisfaction: No Discussions $M = 5.57$, Trial Discussions $M = 5.55$, all $t_s < 1$, ns).

Table 3 presents the results for No Discussions and Trial Discussions cases, respectively. The overall disagreement rate does not differ by condition. Disagreements also did not differ as a function of the complexity of the case, suggesting that the disagreements were not produced by jury misunderstanding of complex evidence.

Given the greater tendency to find for the plaintiff in Trial Discussions cases in Pima County, we analyzed judge–jury agreement rates for Pima County separately. We wanted to know whether or not allowing the Pima County jurors to discuss evidence had created more verdicts that judges rated as inconsistent with the evidence. Judges' agreement with the Pima County verdicts was nearly identical in the two conditions (88% in the No Discussions cases, and 86% in the Trial Discussions cases). Judicial ratings of surprise and satisfaction with the Pima County jury verdicts also did not differ between the two conditions (both $t_s < 1$).

The judicial rating of evidence allowed us to take a further look at the tendency to favor the plaintiff in Pima County Trial Discussions juries. We undertook two sets of logistic regression analyses, the first using whether jurors were permitted to discuss the evidence and the second employing whether they actually engaged in trial discussions, of all cases together and the Pima/Non-Pima Counties separately (see Tables 4 and 5). In the first analysis, which is illustrated in Table 4, we attempted to predict the likelihood of a defense verdict as a function of whether jurors were permitted to discuss the evidence, while assessing the effect of other important variables. Predictor variables included whether jurors engaged in trial discussions,

Table 4. Impact of Trial Discussions Random Assignment and Other Factors on Jury Verdicts: Logistic Regression Results

Variable	<i>B</i>	Wald
Judicial rating of evidence (all cases, <i>N</i> = 161)	.71	17.41**
Non-Pima County cases (<i>n</i> = 99)	.56	7.99**
Pima County cases (<i>n</i> = 62)	.92	7.17**
Trial discussions (all cases, <i>N</i> = 161)	-.93	5.26*
Non-Pima County cases (<i>n</i> = 99)	-.42	.68
Pima County cases (<i>n</i> = 62)	-1.75	5.82*
Case complexity scale (all cases, <i>N</i> = 161)	-.032	.23
Non-Pima County cases (<i>n</i> = 99)	-.09	1.34
Pima County cases (<i>n</i> = 62)	.17	1.24
Judicial rating of attorneys (all cases, <i>N</i> = 161)	-.02	.02
Non-Pima County cases (<i>n</i> = 99)	-.02	.01
Pima County cases (<i>n</i> = 62)	-.19	.42

p* < .05; *p* < .01.

the judicial rating of the evidence (higher numbers indicated that judges saw the evidence as more favorable to the defense), the Case Complexity Scale, and the judicial rating of the relative skills of the attorneys. In the second analysis, illustrated in Table 5, we examined the effect of whether jurors actually engaged in trial discussions along with the other case variables.

In all three analyses (all cases combined, and Pima and non-Pima cases separately), the judicial rating of the evidence has the strongest relationship to the jury's verdict. Case complexity and the rating of attorney skill are never significant predictors of the verdict. The impact of Trial Discussions varies. In the overall analysis and in Pima County, the factor of Trial Discussions decreases the likelihood of a defense verdict. When controlling for whether jurors actually engaged in trial discussions, however, this effect becomes only marginal for both the overall and Pima County verdicts. In the non-Pima County cases, there is no effect of Trial Discussions on the verdict.

Table 5. Impact of Engaging in Jury Discussions and Other Factors on Jury Verdicts: Logistic Regression Results

Variable	<i>B</i>	Wald
Judicial rating of the evidence (all cases, <i>N</i> = 161)	.71	17.95**
Non-Pima County cases (<i>n</i> = 99)	.57	7.97**
Pima County cases (<i>n</i> = 62)	.85	6.72**
Engaged in formal trial discussions (all cases, <i>N</i> = 161)	-.77	3.42
Non-Pima County cases (<i>n</i> = 99)	-.36	.50
Pima county cases (<i>n</i> = 62)	-1.69	3.62
Case complexity scale (all cases, <i>N</i> = 161)	-.01	.02
Non-Pima County cases (<i>n</i> = 99)	-.08	1.15
Pima Country cases (<i>n</i> = 62)	.20	1.91
Judicial rating of attorneys (all cases, <i>N</i> = 161)	-.03	.04
Non-Pima County cases (<i>n</i> = 99)	-.04	.05
Pima County cases (<i>n</i> = 62)	-.11	.12

p* < .05; *p* < .01.

Interactions Among Jurors

Jurors often report that they want to talk about the case and are frustrated that they are not able to do so (Hans & Vidmar, 1986). Without a specific rule providing for trial discussions, as in Arizona, some jurors apparently violate the admonition against discussing the case either with other jurors or with family and friends (Lakamp, 1998; Loftus & Leber, 1986). With this in mind, we hypothesized that providing jurors the opportunity to discuss the case formally within the Rule would reduce illegitimate conversations among small groups of jurors or with external sources.

We found that Trial Discussions jurors were more likely than the No Discussions jurors to say that they had spoken informally with other jurors. In contrast, No Discussions jurors were slightly, albeit not significantly, more likely than Trial Discussions jurors to admit that they had spoken to family or friends. The effect on discussions with family or friends was more pronounced in Pima County. Twenty-three percent (23%) of the No Discussions jurors said they had spoken about the case with friends or family compared to 14% of the Trial Discussions jurors. In a two-way ANOVA, with Pima County and Trial Discussions as factors, both main effects were significant, Pima $F(1, 158) = 17.77, p < .001$, Trial Discussions $F(1, 158) = 3.98, p < .05$, and there was a marginal ($p = .08$) interaction.

With respect to the reported frequency of informal discussions among jurors, it is possible that some Trial Discussion jurors understood the question to include the frequency of juror discussions in which one or more jurors were temporarily absent from the jury room, rather than disapproved discussions among small cliques of jurors. Trial Discussion juries who reported that jurors were always present for discussions (88%) were significantly less likely to engage in informal discussions ($M = 31%$) than juries who report that jurors were present only "most of the time" ($M = 69%$) or "some of the time" ($M = 67%$); $F(2, 55) = 7.60, p = .001$. The Arizona rule permitting juror discussions does not countenance discussions in which individual jurors are not present; technically speaking, the temporary absence of jurors violates the judge's instructions concerning trial discussions. Nevertheless, it would have been useful to have been able to differentiate these types of discussions from informal discussions among small subgroups of jurors.

Two points are worth noting in addition to the difference between conditions. First, by their own admission jurors engage in a relatively high level of informal discussion with family, friends, and other jurors. Fifty-six percent (56%) of the juries included at least one member who admitted having informal discussion about the issues and evidence during the trial with family or friends. Second, No Discussions juries split their informal discussions relatively evenly between jurors and outsiders, whereas Trial Discussions jurors were much more likely to engage in discussions with other jurors.

The higher level of informal discussions among Trial Discussions jurors appears to be a direct result of engaging in formal trial discussions (see Table 3). Analyses with the three discussion groups (No Discussions juries, Trial Discussions juries who did not engage in discussions, and Trial Discussions juries who engaged in discussions) showed that jurors who engaged in formal discussions were more likely than the other two groups to report engaging in informal discussions, perhaps linked to or a carryover from the formal discussions, $F(2, 156) = 12.54, p < .001$.

Conflict and Unanimity

The Arizona Committee argued that the opportunity to discuss the case together could make the jury as a whole more cohesive. Alternatively, jurors who cannot discuss the evidence may use their time together to discuss mutual interests and to develop personal relationships, making jury duty a more pleasant social experience. Examining these alternatives, we find evidence of more conflict and less unanimity by the Trial Discussions juries as compared to the No Discussions juries. In a linear regression analysis, $F(3, 147) = 15.18, p < .0001, R \text{ square} = .24$, both case complexity ($\beta = .16, p = .03$) and the degree to which one or two jurors tended to dominate the deliberations ($\beta = .43, p < .001$) were significant predictors of reported conflict. Trial Discussions was marginal ($\beta = .13, p = .08$).

Arizona does not require unanimity in civil jury verdicts, but instead requires that a supermajority (3/4 of jurors) agree on the verdict (Arizona Rule of Civil Procedure 49(a)). Whether or not the jurors reported being unanimous on the last ballot, however, was associated only with the degree of dominance by one or two jurors (in a logistic regression, B for Juror Dominance = $-.47$, Wald = $5.08, p = .02$) and judges' ratings of the degree to which the evidence favored one side or another (B for judicial rating of strength of evidence = $.38$, Wald = $7.74, p = .005$). Cases in which judges rated the evidence as favoring the defense were more likely to result in unanimous final ballots (54% unanimity), compared to those in which the evidence was evenly split (37% unanimity) or in which the evidence favored the plaintiff (29% unanimity).

However, other measures of the quality of juror interaction produced no significant differences, including the friendliness of the jury, the extent to which jurors reported relying on other jurors, and the degree to which the final jury deliberations represented all points of view or were dominated by one or two people. Jurors had quite favorable views of their fellow jurors and the jury experience (for example, the average rating of the jury's friendliness, intended as a measure of closeness, was 5.8 on a 7-point scale), but these favorable views were not affected by the opportunity to have early discussions about the evidence. Jurors' reports of their own participation and influence in the final jury deliberations were comparable, supporting the overall finding that engaging in trial discussions did not alter these dimensions of interpersonal interaction, at least in a way that was detectable across a range of cases. Deliberation length was affected by the trial length and the degree of conflict among jurors, but not by the opportunity to discuss evidence early in the trial as had been hypothesized. Nor were Trial Discussions jurors quicker to focus on critical issues during deliberation.

DISCUSSION

The results of the field experiment fulfill neither the fondest hopes nor the worst nightmares of supporters and critics of the trial discussions jury reform. First, let us consider what the evidence has shown about the potentially beneficial effects of the reform. The research literature on the value of group decision making for

error correction suggests that interim discussions could be advantageous to jury comprehension of evidence and instructions. In our study, jurors who have themselves engaged in trial discussions report that the discussions are very helpful in resolving confusion about the evidence. Yet this assistance, if it exists, does not translate into observable differences in jurors' reports of their understanding of evidence and law, or differences in agreement with the judge over the proper verdict. The judge's ratings of which side the evidence favored largely agreed with jury verdicts, and they did not differ significantly between the experimental and control conditions. Furthermore, judges had very favorable views of jury understanding, believing that jurors understood the key issues and evidence in the cases before them.

Given that jurors themselves reported that the discussions helped them better understand the evidence, we would like to comment on the study's lack of statistically significant results pertaining to these reported benefits. First, there may actually be no benefits, and jurors are simply mistaken about the role of trial discussions in helping them better understand the evidence. Second, there could be some benefits, but our research methodology and measures were not sensitive enough to pick them up. We relied upon general questions posed to jurors and other trial court participants, rather than assessing juror understanding more directly as other methodologies such as mock jury research are able to do. A general comprehension instrument that would have been applicable to all civil juries was not feasible. Jurors' views of their own ability to understand the law and evidence are quite positive in both conditions, and judges report that jurors appear to understand the key issues in their cases in virtually all trials. Furthermore, to reliably affect verdicts and judge-jury agreement rates, any improvements would have to be key, substantial, and present in a reasonable number of cases. Heuer and Penrod (1996) and Lempert (1975) make similar observations about the challenge in field experiments to reliably detect improvements in juror comprehension as a result of jury reforms. Other studies such as mock jury laboratory experiments which control evidence and can test comprehension for relevant issues in a trial and observe errors in deliberations should be better able to address this question.

Trial discussions were also predicted to increase prejudgment. Here, the findings of our study indicate that being able to discuss evidence with other jurors during the trial lessens the degree of uncertainty that jurors feel about the case at the start of deliberation. Trial Discussions juries were significantly less likely to report that they were unsure about which side they favored as final deliberations commenced, compared to No Discussions juries. A change in the percentage of jurors who are unsure at the beginning of deliberations is a plausible effect of allowing jurors to engage in discussions about evidence during the trial. Trial discussions allow jurors to air their impressions of the evidence as it is presented and to receive the input of other jurors. Nevertheless, some other variables that we expected would differ if jurors were prejudging the case (when jurors report leaning toward one side or the other or making up their minds; their evaluations of plaintiff and defense attorneys, and the rated importance of the first and last witnesses) did not vary as a function of trial discussions.

Did the opportunity to discuss the case with other jurors produce a plaintiff

tilt? In Pima County, by chance, the cases randomly assigned to the Trial Discussions conditions were more likely to be rated by the judges as stronger for the plaintiff, and juries were more likely to decide in favor of the plaintiff in the Trial Discussions condition. We attempted to control for these case differences in evidence in logistic regression analyses. These analyses show that when Trial Discussions and the evidence ratings are considered simultaneously as predictors of the jury's verdict, Trial Discussions has a significant effect (although actually engaging in trial discussions has only a marginal effect). We have no compelling explanation for why Trial Discussions should increase plaintiff verdicts in Pima County while not influencing plaintiff verdicts in the other Arizona counties we studied. We observe that the weaker effect for engaging in discussions suggests that case differences rather than the impact of discussions may be driving this effect. Given our lack of success in achieving comparable distribution of cases across conditions in Pima County, we urge caution in the interpretation of this finding.

One of the advantages of random assignment is the ability to make causal statements about the impact of an independent variable. However, in a field experiment like this one, by chance or by differential noncompliance, case characteristics can vary between the conditions, as we discovered with case complexity and judicial assessments of the evidence. Currently, a second experiment of the impact of trial discussions is underway that may provide a more thorough understanding of how these discussions affect the dynamics of jury decision making and juror interaction.

Some commentators such as MacCoun (1995) warned that trial discussions might lead to more shared biases among jurors, since they would be privy to general views of the evidence held by other jurors and might adopt those views early on. Our methodological approach was limited in its ability to detect such biases, if they exist. However, some of our results suggest the opposite—that trial discussions promote more vigorous and contentious debate. Trial Discussions jurors reported greater conflict and less unanimity on the final ballot. This is reminiscent of the “verdict-driven” juries identified by Hastie et al. (1983). Nevertheless, jurors in both conditions expressed the same high level of satisfaction with the jury's result, despite the reports about greater conflict in Trial Discussions cases. Arizona's law permitting less than unanimous civil verdicts, and the opportunity for civil jurors to bargain among themselves over the size of an award, probably helps to explain why substantial nonunanimity did not translate into hung juries. However, the findings suggest that the reform's impact could be different in states with a unanimity requirement. Its relevance in criminal cases, which require the prosecution to meet a higher standard of proof than plaintiffs in civil cases, is also unknown.

This paper also reports for the first time systematic data about the extent to which jurors engage in discussions with family, friends, and other jurors outside the officially sanctioned trial discussions. As some have suspected, the extent of these informal discussions is substantial. Despite judicial admonitions not to discuss the case, over half the juries included at least one member who admitted he or she had spoken to friends or family about the issues and evidence during the trial. The opportunity to engage in trial discussions tends to shift some of these talks with friends and family to the other jurors, but it also substantially increases the degree to which jurors talk informally with one another about the case. More detailed

information about the frequency and content of these informal discussions is needed to make a fully informed judgment about whether these informal discussions cross the line into prohibited topics of conversation. We hope the follow-up study to this experiment will provide this information.

This field study poses some interesting questions for both judicial policy makers and social scientists beyond the immediate implications of permitting jurors to discuss the evidence before final deliberations. Arizona has been at the forefront of the country's jury reform movement and has spearheaded the development of the new active juror approach to decision making based on scholarly research from the fields of communications, educational psychology, and jury decision making. This approach is a marked shift from the adversarial assumption of the traditional passive juror model. As this first test of this model in the field, this study highlights the value of using multiple methodologies to investigate the impact of the active juror approach. We hope that these findings will prompt others to investigate the questions raised in this study using more traditional laboratory methodologies to isolate and measure opinion change and the impact of group discussions.

ACKNOWLEDGMENTS

The research reported in this article was funded by State Justice Institute grant SJI-96-12A-B-181 to the National Center for State Courts. In addition, a fellowship in the Center for Advanced Study at the University of Delaware facilitated Valerie Hans's work on this project. The authors thank Chief Justice Thomas A. Zlaket, the incredibly cooperative Arizona judiciary, and the members of the project's Advisory Committee for their valuable suggestions about the design and execution of the study. The Advisory Committee included Hon. B. Michael Dann, Shari Seidman Diamond, Hon. Judith McConnell, Hon. Judith E. Retchin, G. Marc Whitehead, and Elizabeth Wiggins. Special thanks are owed to Hon. B. Michael Dann and George Logan, Project Director, Arizona Supreme Court, for facilitating the implementation of the administrative rules and procedures needed to conduct this field experiment, and to Prof. Richard Lempert for comments on an earlier draft of this article. The points of view expressed are those of the authors and do not necessarily represent the official positions or policies of the State Justice Institute or the National Center for State Courts.

REFERENCES

- Arizona Rules of Civil Procedure Rule 39(f), Rule 49(a).
 Arizona Supreme Court Order No. R-92-0004 (1995).
 Arizona Supreme Court Order No. R-94-0031 (1995).
 Arizona Supreme Court Committee on More Effective Use of Juries. (1994, November). *Jurors: The power of twelve*. Phoenix, AZ: Author.
 Commonwealth v. Kerpan, 498 A.2d 829 (Pa. 1985).
 Dann, B. M. (1993). "Learning lessons" and "speaking rights:" Creating educated and democratic juries. *Indiana Law Journal*, 68, 1229-1279.
 Diamond, S. S. (1997). Illuminations and shadows from jury simulations. *Law and Human Behavior*, 21, 561-572.

- Diamond, S. S., & Casper, J. D. (1992). Blindfolding the jury to verdict consequences: Damages, experts, and the civil jury. *Law and Society Review*, 26, 513–563.
- Ellsworth, P. C. (1989). Are twelve heads better than one? *Law and Contemporary Problems*, 52, 205–224.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York: Random House.
- Hans, V. P., Hannaford, P. L., & Munsterman, G. T. (1999). The Arizona jury reform permitting civil jury trial discussions: The views of trial participants, judges, and jurors. *University of Michigan Journal of Law Reform*, 32, 349–377.
- Hans, V. P., & Vidmar, N. (1986). *Judging the jury*. New York: Plenum, Press.
- Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, 93, 258–268.
- Hastie, R., Penrod, S. D., & Pennington, N. (1983). *Inside the jury*. Cambridge, MA: Harvard University Press.
- Heuer, L., & Penrod, S. (1989). A field experiment with written and preliminary instructions. *Law and Human Behavior*, 13, 409–430.
- Heuer, L., & Penrod, S. (1996). Increasing juror participation in trials through note taking and question asking. *Judicature*, 79, 256–262.
- Hunt v. Methodist Hosp., 485 N.W.2d 737, 743 (Neb. 1992).
- Kalven, H., Jr., & Zeisel, H. (1966). *The American jury*. Boston: Little, Brown.
- Kassin, S. M., & Wrightsman, L. S. (1979). On the requirements of proof: The timing of judicial instruction and mock juror verdicts. *Journal of Personality and Social Psychology*, 37, 1877–1887.
- Kassin, S. M., & Wrightsman, L. S. (1988). *The American jury on trial: Psychological perspectives*. New York: Hemisphere.
- Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996). Bias in judgment: Comparing individuals and groups. *Psychological Review*, 103, 687–719.
- Lakamp, N. K. (1998). Deliberating juror predeliberation discussions: Should California follow the Arizona model? *U.C.L.A. Law Review*, 45, 845–879.
- Landsman, S. (1984). *The adversary system: A description and defense*. Washington, DC: American Enterprise Institute for Public Policy Research.
- Lempert, R. O. (1975). Uncovering “non-discernible differences:” Empirical research and the jury size cases. *Michigan Law Review*, 73, 643–708.
- Lind, E. A., & Tyler, T. R. (1998). *The social psychology of procedural justice*. New York: Plenum Press.
- Loftus, E. F., & Leber, D. (1986). Do jurors talk? *Trial*, 1986, 59–60.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and polarization: The effect of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37, 2098–2109.
- MacCoun, R. (1995, July). *Improving jury comprehension in criminal and civil trials*. RAND Testimony Series No. CT-136.
- Meggs v. Fair, 621 F.2d 460 (1st Cir. 1980).
- Munsterman, G. T. (1996). A brief history of state jury reform efforts. *Judicature*, 79, 216–219.
- Munsterman, G. T., Hannaford, P. L., & Whitehead, G. M. (eds.). (1997). *Jury trial innovations*. Williamsburg, Virginia: National Center for State Courts.
- Myers, R. D., & Griller, G. M. (1997). Educating jurors means better trials: Jury reform in Arizona. *Judges’ Journal*, 36, 13–17, 51.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of human judgment*. Englewood Cliffs, NJ: Prentice Hall.
- Pearson, R. W., Ross, M., & Dawes, R. M. (1992). Personal recall and the limits of retrospective questions in surveys, in J. M. Tanur (Ed.), *Questions about questions: Inquiries in the cognitive bases of surveys* (pp. 65–94). New York: Russell Sage Foundation.
- Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, 51, 242–258.
- Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology*, 62, 189–206.
- People v. Blondia, 245 N.W.2d 130 (Mich. App. 1976).
- People v. Hunter, 121 N.W.2d 442 (Mich. 1963).
- People v. Monroe, 270 N.W.2d 655 (Mich. 1978).
- Saks, M. J. (1981). *Small-group decision making and complex information tasks*. Washington, DC: Federal Judicial Center.
- Shtabsky, J. E. (1996). Comment. A more active jury: Has Arizona set the standard for reform with its new jury rules? *Arizona State Law Journal*, 28, 1009–1033.
- State v. Washington, 438 A.2d 1144 (Conn. 1980).
- United States v. Meester, 762 F.2d 867 (11th Cir. 1985).

- United States v. Wexler, 657 F.Supp. 966 (E.D. Pa. 1987), rev'd en banc 838 F.2d 88 (3rd Cir. 1988).
- Visher, C. (1987). Juror decision making: the importance of evidence. *Law and Human Behavior*, 1987, *11*, 1-17.
- Weld, H. P., & Danzig, E. R. (1940). A study of the way in which a verdict is reached by a jury. *American Journal of Psychology*, 1940, 518-537.
- Wilson v. State, 242 A.2d 194 (Md. 1968).
- Winebrenner v. United States, 147 F.2d 322 (8th Cir. 1945).