

1995


Are Twelve Heads Better Than One?

Phoebe C. Ellsworth

University of Michigan Law School, pce@umich.edu

Available at: <https://repository.law.umich.edu/articles/741>

Follow this and additional works at: <https://repository.law.umich.edu/articles>

 Part of the [Courts Commons](#), [Criminal Procedure Commons](#), [Law and Psychology Commons](#), [Law and Society Commons](#), and the [Litigation Commons](#)

Recommended Citation

Ellsworth, Phoebe C. "Are Twelve Heads Better Than One?" *Law Quad. Notes* 38, no. 2 (1995): 56-64.

This Article is brought to you for free and open access by the Faculty Scholarship at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Articles by an authorized administrator of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.

This article has been condensed with permission from one originally published in Law and Contemporary Problems, Vol. 52, No. 4. Most reference citations have been deleted. For a fully-footnoted copy of the original article, please contact the LQN editor.

Are

— BY PHOEBE C. ELLSWORTH

Few advocates of the jury system would argue that the average juror is as competent a tribunal as the average judge. Whatever competence the jury has is a function of two of its attributes:

**its number
and its
interaction.**

The fact that a jury must be composed of at least six people, with different backgrounds, experiences, and perspectives, provides protection against decisions based on an idiosyncratic view of the facts. In addition, the jury must be chosen in a manner that reflects a representative cross-section of community opinion. The jury's competence, unlike that of the judge, rests partly on its ability to reflect the perspectives, experiences, and values of the ordinary people in the community — not just the most common or typical community perspective, but the whole range of viewpoints.

Representativeness is important not only for ensuring “the essential nature of the jury as a tribunal embodying a broad democratic ideal,”¹ but because it affects the jury's competence directly. Failure to assure that any given group has a fair chance of participation “deprives the jury of a perspective on human events that may have unsuspected importance in any case that may be presented.”²

A jury decision, however, is more than an average of the verdict preferences of six or twelve citizens who represent a variety of experiences. Ideally, the knowledge, perspectives, and memories of the individual members are compared and combined, and individual errors and biases are discovered and discarded, so that the final verdict is forged from a shared understanding of the case. This understanding is more complete and more accurate than any of the separate versions that contributed to it, or indeed than their average. This transcendent understanding is the putative benefit of the deliberation process.

12 HEADS



better than

If it does nothing else, group deliberation (except in extraordinarily one-sided cases) forces people to realize that there are different ways of interpreting the same facts. While this rarely provokes a prompt revision of their own views, it necessarily reminds the jury members that their perceptions are partly conjectural — an obvious truth, but one that is otherwise unlikely to occur to them.

A judge does not have this vivid reminder that alternative construals are possible. A judge, however, has experience on the bench and training in the law. Critics of the jury often focus on the incompetence of people chosen as jurors, compared to that of the judge. At best, the venire consists of a representative sample of the community, with a few members having genuine expertise, a large number who are simply average citizens, and a few others who are distinctly below average. In practice, many of the better-educated jurors are excused from service, and others who show knowledge or ability relevant to the particular case at trial may be challenged

during the voir dire. Attorneys sometimes select jurors for incompetence.³ Thus, some have argued that the average jury is not only less competent than the average judge, but is also less competent than a random sample of twelve citizens from the community.

Historically, the debate over the competence of juries has been less than enlightening. In particular, there are two conspicuous omissions.

First, there is a great reluctance to define competent decision-making. Social scientists who turn to the legal literature in search of criteria by which to evaluate the jury are likely to find it a frustrating experience. It is extremely difficult to design research that will contribute useful information to the debate on competence when the concept of competence is not defined.

Second, most of the social science research and much of the legal debate has focused primarily on the jury's verdict, an extremely crude measure of competence, and one that tells us very little about what juries actually do.

One way to look at jury functioning is to break down the jury's task into components, and look at the way the jury deals with each one. Pennington and Hastie⁴ have provided a useful list:

1) The jury members must "encode" the information they get at trial. A competent jury must pay attention to the testimony and remember it.

2) The jury must define the legal categories. A competent jury should define these categories as they are presented in the judges' instructions.



1. Ellsworth & Getman, "Social Science in Legal Decision-Making," in *Law and the Social Sciences* 596 (L. Lipson & S. Wheeler eds. 1986).
2. *Peters v. Kiff*, 407 U.S. 493, 503-04 (1972) (plurality opinion of Marshall, J., joined by Douglas and Stewart, JJ.)
3. See J. Van Dyke, *Jury Selection Procedures* (1977); V. Hans & N. Vidmar, *Judging the Jury* 63-78 (1986).

3) The jury must select the admissible evidence and ignore evidence that is inadmissible.

4) The jury must construct the sequence of events.

5) The jury must evaluate the credibility of the witnesses.

6) The jury must evaluate the evidence in relation to the legal categories provided in the instructions. That is, certain elements of the story the jury constructs are particularly important in determining the appropriate verdict. The jury must identify these elements and understand how differences in the interpretation of the facts translate into differences in the appropriate verdict choice.

7) The jury must test its interpretation of the facts and the implied verdict choice against the standard of proof: preponderance of evidence, clear and convincing evidence, or beyond a reasonable doubt.

8) The jury must decide on the verdict.

In discussing my research on jury deliberations, I present data and some impressions of how the jury performs these tasks; I also discuss some other aspects of jury deliberation.

The research itself involved close analysis of eighteen mock juries in the first hour of deliberation. Because of the small sample size, statistical analysis of the data generally would be misguided. The study is most usefully considered as an intensive case study of the process of jury deliberation. However, the fact that there are eighteen cases rather than one makes it considerably more useful than the usual case study, because it allows for some assessment of the variability of juries exposed to the same stimulus.

A major drawback is that none of the juries reached a verdict in the hour allotted to them. Thus, the study is most useful as an exploration of how juries structure their task, how well they deal with the facts and the law, and what things they discuss. It is very likely that at some point juries move into an "endgame" that may differ substantially from the phases preceding it.



Ideally, the knowledge, perspectives, and memories of the individual members are compared and combined, and individual errors and biases are discovered and discarded, so that the final verdict is forged from a shared understanding of the case.

Method

Two hundred and sixteen adults eligible for jury service in Santa Clara or San Mateo County, California, participated in the deliberation study and provided usable data. Thirty-three of them were recruited from the venire lists of the Santa Clara County Superior Court after completing their terms as jurors. The remainder had responded to a classified advertisement in local newspapers asking for volunteers for a study of "how jurors make decisions," or were referred by friends aware of or participating in the study. Each subject was paid ten dollars for participation.

The sample was fairly representative of the suburban upper-middle-class community surrounding Stanford University, except that males and minorities were underrepresented. The sample was 93 percent white and 65.3 percent female. The average age of the subjects was 43, and 63 percent of the sample was employed outside the home. The median educational level was slightly less than a college degree. Finally, 46 percent of the sample had previously performed jury duty, while 37 percent had actually served on juries.

Subjects watched a videotape of a simulated homicide trial that represented all major aspects of an actual criminal trial. After hearing the evidence, arguments, and instructions, the jurors gave an initial verdict. Jurors were then assigned to twelve-person juries and allowed to deliberate for one hour.

We chose to use a videotape prepared by Reid Hastie for use in his research on jury unanimity.⁵ This tape is representative of the procedures, setting, style, and issues that commonly occur in actual homicide trials. The case was complex enough to afford several plausible interpretations and verdict preferences. It resembled most real murder trials in that there was no question that the defendant had killed the victim; rather, the evidence centered on the precise sequence of events preceding the killing and on the defendant's state of mind at the time. Finally, the tape was far more vivid and realistic than any other simu-

lated trial materials we have encountered. It was highly unlikely that we could have constructed a better tape with our resources.

Hastie's videotape is a reenactment of an actual homicide case based on a complete transcript of the original trial, with a judge and experienced criminal attorneys playing roles based on the actual judge's instructions and lawyers' arguments. We modified the tape in two ways for the present research. First, we shortened it slightly by deleting one defense witness whose testimony added little. Second, we replaced the segment of the original tape containing the original instructions, which had been based on Massachusetts law, with a new sequence in which the applicable California law was given. Pretesting indicated that the tape was regarded as convincing and realistic.

In the trial videotape, the defendant, Frank Johnson, is charged with first-degree murder for the stabbing of Alan Caldwell outside a neighborhood bar. The prosecution brings evidence that the defendant and victim had argued in the bar earlier that day, and that Caldwell had threatened the defendant with a straight razor. Johnson had left after the argument, but had returned with a friend that evening. Caldwell later came into the bar, and he and the defendant went outside and began to argue loudly. Two witnesses testify that they saw Johnson stab down into Caldwell's body. The victim's razor was subsequently found, folded, in his left rear pocket.

For the defense, Johnson testifies that he had returned to the bar that evening on the invitation of his friend and had entered only after ascertaining that Caldwell was not there. Caldwell had come in later and had asked Johnson to step outside, presumably for the purpose of patching up their quarrel. Once outside, Caldwell had hit him and come at him with a razor. Johnson had pulled out a fishing knife which he often carried in his pocket and Caldwell had run onto the knife. In cross-examination, the defense attorney cast doubt on the ability of the prosecution's eye witness to see the scuffle, and showed that medical evidence cannot establish whether the defendant stabbed down into the victim or whether the victim ran onto the knife.

Four verdicts are possible in this case, depending upon the jury's findings of facts. The defendant may be guilty of first-degree murder, of second-degree murder, or of voluntary manslaughter, or he may be not guilty for reason of self-defense or accidental homicide.

The study was conducted on weekend afternoons at Stanford University. Each subject group consisted of twelve to thirty-six subjects. Upon arrival, all subjects were given a brief overview of the study and asked to fill out an informed consent form and a preliminary questionnaire focusing on demographic characteristics, general attitudes toward the death penalty and toward criminal defendants, and general attitudes with respect to crime control and due process. The experimenter then introduced the videotape and instructed subjects to pay close attention because afterwards, they would be asked to deliberate to reach a verdict based on the facts of the case and the judge's instructions, just as if they were actual jurors.

Deliberations

As soon as the videotape was over, the experimenter asked the subjects to indicate their verdict preferences on an initial verdict questionnaire by checking one of four choices: first-degree murder, second-degree murder, manslaughter, or not guilty. After collecting the questionnaires, the experimenter announced assignments to jury panels and directed each jury to a separate room for deliberation. These were seminar rooms equipped with a long table and a video camera and two ceiling microphones to record deliberations for later analysis. The equipment also allowed the experimenters to view the deliberations on a monitor outside the room, in order to detect problems that might jeopardize the validity of the study.

Once the subjects were settled in the jury room, the experimenter told them that their next task was to discuss the case and try to reach a verdict. They were assured that their immediate postvideotape verdict was confidential

and that they need not feel committed to it. They were also told that most juries begin by taking a straw vote, and that in any case they should choose a foreman before beginning their deliberation. The experimenter continued as follows:

As you discuss the case, it is important to put yourselves into the role of jurors. Imagine that you are a real jury and that your verdict will actually determine the fate of the defendant you saw on the tape. We want you to make your decision only on the basis of what you saw on the tape. Although the characters in the trial you saw were actors, we want you to treat them as if they were real. In short, we want you to make the decision you would make if you were a real jury and if you had seen in court exactly what you saw on the tape.

The experimenter closed by informing the subjects that they had one hour in which to deliberate, and that they should try to reach a decision in that time, although quite possibly one hour would not be long enough to reach a consensus. The purpose of this instruction was simply to assure that the subjects worked on their deliberation seriously and tried to reconcile their differences of opinion. We did not ask them to take a vote at the end of the hour, and we did not expect them to reach a verdict.

Subjects were then left to discuss the case. Although they appeared to be slightly self-conscious in the presence of the recording equipment for the first minute or two, the jurors became highly involved in the discussion and seemed to forget about the camera as soon as the deliberations revealed disagreements

4. Pennington & Hastie, "Juror Decision Making Models: The Generalization Gap," 89 *Psychological Bulletin* 246, 249-55 (1981).
5. R. Hastie, S. Penrod, & N. Pennington, *Inside the Jury* (1983).

among the members, which occurred almost immediately for each jury. After an hour the experimenter returned, stopped the deliberation, and handed out the postexperiment questionnaires.

The videotaped jury deliberations were transcribed, and the transcripts were divided into units. In devising the coding scheme, I identified thirty major issues in the case. A unit, by definition, could contain no more than one of these issues. Short utterances occasionally contained none; long utterances were divided into units corresponding to the number of issues. Each transcript was coded by one or more of three trained coders. Coders were given lists of 100 case facts, 18 major issues, and 60 legal instructions; at various points, two coders were asked to code the same jury in order to calculate inter-coder reliability. Each unit was coded for the general nature of the statement (issue, fact, law, vote, procedural comment, and so on), correctness, pro defense or pro prosecution position, and the particular fact, issue, or point of law that was mentioned. Coders met weekly with me to resolve questions and settle differences.

Choosing a foreman

All juries began by choosing a foreman, not surprisingly, since the experimenter had instructed them to do so. The foreman was always chosen very quickly, with a minimum of discussion.

The process of foreman selection can be summed up by the phrase “choose a man who says he has experience.”

Although 65 percent of the jurors were female, sixteen of the eighteen foremen were male.⁶ On the jury composed of eleven women and one man, the man was chosen. When the jurors had arrived in the room and settled in their seats, someone would point out that their first job was to choose a foreman, and then typically someone would ask, “Has anybody had any experience with this sort of thing?” A man would claim experience, and the other jurors would agree that he should take the job. Occasionally two men would claim experience and a brief “after you, Alphonse” discussion would ensue until one of them said, “all right, I’ll do it.” These two scenarios account for foreman selection in ten of the eighteen juries.

Since we knew which of our subjects had actually served on real juries, we were able to find out whether the people chosen as foreman were actually more likely to have had prior jury experience than the other jurors. They were not more experienced: 39 percent of the foremen had served on juries, as compared with 36 percent of the other jurors, an insignificant difference. Thus, a foreman is someone who claims experience, not necessarily someone who has it.

On the remaining eight juries, five foremen (four male, one female) were chosen because they were sitting in one of the seats on the ends of the table, and three (two male, one female) were individuals who had opened the discussion by volunteering for the position. Altogether, nine of the foremen were sitting at the head of the table, and four others were sitting in the chair right next to the head. Table position is by no means a subtle proxemic cue that exerts an unconscious influence on the jurors; in the majority of cases the jurors explicitly gave table position as their reason for their choice — “you should do it, you’re sitting in the right place.”

These data suggest that jurors give little consideration to their selection of foremen. They are generally given no information on what qualifications to look for, so they have little to guide them but their background knowledge and stereotypes of the jury, gained from the media and other sources. In the movies, the foreman sits at the head of the table.

In addition, at the time that the foreman is chosen, most jurors may still regard their task as a relatively simple one, because the extent of disagreement on the jury has not yet been revealed. They may not think it makes much difference who is chosen foreman, because they see the case as straightforward and do not anticipate serious disputes. Finally, since no disagreements have yet been revealed, it is likely that strong norms of courtesy prevail at the time that the foreman is selected. Once someone has been suggested, the others may think it is impolite to question his or her ability.

Taking the task seriously

Once the foreman was selected, the juries took one of two approaches to their task. One-half of the juries began by taking a vote, roughly evenly divided among show-of-hands, secret ballot, and a go-around procedure in which each juror states a position and says a little about his or her reasons for taking that position. The other half of the juries began by discussing the facts and issues in the case. The judge’s instructions contained a caution to the jurors not to become unduly committed to their position but to remain open-minded. A few jurors interpreted these instructions to mean that they should not begin deliberations with a vote.

Hastie and his colleagues⁷ have proposed that when a jury postpones a formal vote, it is freer to raise issues and discuss them open-mindedly. When a jury begins by voting, people feel committed to the position they have publicly expressed, and spend their time defending their position rather than trying to understand the facts and the law. Our data generally support Hastie’s findings. Juries that postponed voting spent more time talking about the important issues in the case, and brought out more facts. One might hypothesize that juries that voted early would spend more time discussing the relevant law, because they

6. This gender bias in choice of a foreperson has changed little over the last forty years. See Strodtbeck, James, & Hawkins, “Social Status in Jury Deliberations,” 22 *Amer. Soc. Rev.* 713 (1957). It occurs not only in mock jury research but in real trials. See Kerr, Harmon & Graves, “Independence of Multiple Verdicts by Jurors and Juries,” 12 *J. Applied Soc. Psychology* 12, 24-25 (1982); Note, “Gender Dynamics and Jury Deliberations,” 96 *Yale Law Journal* 593 (1987).

7. See Hastie, Penrod, & Pennington, *supra* note 5; see also Hawkins, “Interaction Rates of Jurors Aligned in Factions,” 27 *Am. Soc. Rev.* 689 (1962).

8. See H. Kalven & H. Zeisel, *The American Jury* 486 (1966).

would need to define the legal verdict categories before they could vote. This, however, was not the case. Early versus late voting did not predict the amount of time spent discussing the law.

Whether or not a jury began by voting, it was quickly apparent to members of the jury that they disagreed about the appropriate verdict. As soon as these disagreements emerged, the character of the deliberation changed. During foreman selection there was an atmosphere of conviviality in the jury room, along with some degree of self-consciousness. A few jurors joked about the videotape camera. Once the discussion or an early vote revealed differences of opinion, there were no more references to the camera and few jokes of any sort. They kept their attention focused on the case.

On the average, 47 percent of their utterances concerned the facts of the case; 32 percent addressed the important contested issues (for example, the defendant's state of mind, provocation, angle of the knife thrust, ability of witnesses to see the crime); 21 percent dealt with the law and the judge's instructions; and 7 percent were votes or discussions about calling for a vote. (A given utterance could involve both a fact and an issue, or a fact and a point of law, so the percentages do not add to 100 percent.) These proportions are quite comparable to those found by Hastie and colleagues, whose juries saw the same case but deliberated to a final verdict. The criticism that juries approach their task in a frivolous manner receives no support from this study or from any other serious empirical research on the jury.

Discussing facts and issues

Whether or not the jury began with a vote, the general progression of the deliberation moved from an emphasis on facts toward an emphasis on law. In juries that did not begin by voting, the initial discussion resembled a random walk through the facts and issues. A topic would be raised, discussed briefly, and replaced by a totally different topic, with little attempt to organize the discussion and no attempt to resolve the issues.



When a jury begins by voting, people feel committed to the position they have publicly expressed, and spend their time defending their position rather than trying to understand the facts and the law.

These juries conformed very closely to Kalven and Zeisel's observation that "the talk moves in small bursts of coherence, shifting from topic to topic with remarkable flexibility. It touches an issue, leaves it, and returns again."⁸ During the hour of deliberation, the important facts and issues would come up again and again, while trivial issues would be dropped, and new issues added. Typically, as an issue was examined and re-examined, there would be movement toward consensus. For example, one of the most important pieces of evidence in the trial was the coroner's statement that he found the victim's razor folded up in his back left pocket. Had the victim been coming at the defendant with the razor, a self-defense scenario would have been very plausible. The defendant and his friend claimed to have seen the razor drawn; two other witnesses testified that they did not see the razor. Most juries raised this issue early and dropped it without fully considering the implications.

In subsequent discussions, someone would raise the possibility that the victim somehow, in a reflex-like action, could have folded up the razor and pocketed it after he was stabbed, or that someone else (the policeman, the ambulance doctor, or a passer-by) might have picked it up and put it in the dead man's pocket. The jury would eventually conclude that these possibilities were farfetched, and agree that the victim never pulled the razor during the fatal confrontation. As a consequence, some juries would reject the possibility of self-defense and a few would turn their attention to the relevant question of the defendant's possible *belief* that the razor was drawn. In general, over the course of deliberation, jurors appear to focus more on the important facts and issues, come to a clearer understanding of them, and approach consensus on the facts.

In juries that began with a vote, the discussion tended to be slightly more organized. The average distribution of verdicts prior to deliberation was one for first-degree murder, two for second-degree murder, six for manslaughter, and two for not guilty. Although none of the juries showed exactly this pattern, most

of them had a majority of votes in the two middle categories with outliers for not guilty or for both not guilty and first-degree murder. A common tactic was for the middle jurors to begin by asking the outliers to explain their deviant position, typically starting with the proponents of first-degree murder. Whether or not the jury began with a vote, however, issues were raised and dropped fairly unsystematically, then raised again; slowly, progress was made. Little by little, most juries resolved the issues of fact and spent an increasing proportion of their time on the central issue: the defendant's state of mind.

Dealing with the facts

Kalven and Zeisel conclude that "the jury does by and large understand the facts and get the case straight."⁹ On the whole, the data from this study support that conclusion. The juries in our study spent more time discussing the facts of the case (47 percent of the units included references to facts brought out in testimony) than anything else. These were rarely purely factual statements. Most of the time facts were raised in connection with a contested issue, a reference to common sense or knowledge, a hypothetical scenario, or a reference to the law.

Most of the juries managed to sort out the factual issues fairly well during the process of deliberation. Conflicting testimony (for example, about the angle of the knife thrust) was recognized as such, so that juries ended up correctly attributing different versions of the story to different witnesses. Questions regarding the distance and angle of vision of the various witnesses were generally resolved correctly, and errors of fact generally were corrected. None of the juries maintained an erroneous perception of an important fact after the hour of deliberation. Implausible suggestions generally were discussed and rejected, as

in the case of someone putting the razor in the victim's pocket after he was stabbed.

Jurors tended to focus on testimony that favored their initial verdict preferences: Testimony about the previous confrontation between the two men was generally raised by jurors who favored a murder verdict, whereas testimony that the victim punched the defendant immediately before the killing was generally raised by jurors who favored manslaughter or self-defense. This tendency is not a weakness, but rather a benefit of the deliberation process — the opportunity it affords for comparing several different interpretations of the events along with the supporting factual evidence.

For most of the juries in this study, discussion of the facts and issues dominated the first part of the hour. Among the juries that voted early, there was usually some discussion of the judge's instructions in order to arrive at the verdict categories, but the discussion was generally quite superficial. During the course of the factual discussions, the central issues of disagreement emerged, and jurors attempted to persuade each other. Agreement on the facts, however, did not lead to substantial agreement on the central issue of the case: the defendant's state of mind. Jurors tried to persuade each other that their construals of the facts made sense. The discussions often became heated, few opinions were changed, and at some point (often, but not necessarily in connection with a vote), the jurors would turn to the legal definitions of the verdict choices for guidance.

Dealing with the law

Juries worked hard to understand the law. They spent an average of 21 percent of their time discussing the judge's instructions. Following the hour of deliberation, jurors were given an eighteen-question true-false test on elements of the judge's instructions. On average, the jurors answered 11.7 of the questions correctly, a result not significantly different from random guessing. On a postdeliberation multiple-choice

test of factual issues, however, jurors performed quite well, answering correctly an average of 8.8 out of 14 questions (since there were four response alternatives, 3.5 correct answers would be expected by chance). These results suggest that the deliberation process works well in correcting errors of fact but not in correcting errors of law.

An examination of the statements jurors made about the law during the course of their deliberations provides further gloomy detail. We coded all statements jurors made about the law as correct, incorrect, or unclear. Remarks were coded as correct even if they were incomplete. For example, the statement "first-degree murder involves premeditation" would be scored as correct. Statements were scored as incorrect if they were unambiguously wrong; for example, "second-degree murder involves premeditation."

Statements that were coded as unclear were usually statements about verdict-evidence relationships; for example, "If Johnson knew that Caldwell would be there, it's premeditation." While this statement is technically false, because returning to a bar knowing that one's enemy is there does not necessarily imply intent to kill, it was scored as unclear, because the juror could have meant that Johnson's knowledge was a relevant consideration in determining premeditation. Thus, we did not code statements as incorrect unless there was no plausibly correct construal.

Given this rather lenient coding, we found that only half of the references to the law (631) were accurate, even when credit was given for partial accuracy. We found that 609 were not correct (28 percent unclear; 21 percent definitely incorrect). Whereas factual errors tended to be corrected during deliberation, errors of law were not corrected. Considering instances where the jury changed its position, 52 percent of them involved replacing an erroneous response with a correct one, and 48 percent involved replacing a correct response with an erroneous one.

9. *Id.* at 149.

These results are quite distressing, since they mean that the jury does not recognize the right answer when it hears it. Juries who have heard the right definition are as likely to reject it as juries who have heard the wrong one. The jury as a whole does not profit from the abilities of its best members when it comes to questions of law.

During the course of deliberation, jurors generally fought to defend their correct opinions of the facts but not their correct versions of the legal standards. Typically the most forcefully expressed position prevailed, whether or not it was correct. Most of the jurors' discussions of substantive law (that is, the definitions of the verdict categories) conveyed an impression of considerable uncertainty ("Was it . . . I think it was something about passion?"), and jurors who seemed confident about the law were often believed, whether or not their statements corresponded to the judge's instructions.

Of the 1,752 units across all juries that referred to the law, only seventy-five (4 percent) were error corrections. Only 12 percent of the 609 incorrect and unclear statements were corrected. Only 10 percent of the 1,285 references to the verdict choices addressed the distinctions between them. Of these, 26 percent were correct statements, 11 percent were definitely incorrect, 42 percent were unclear, and 21 percent were questions. Examining each jury's last definition of the four verdict choices during the course of the hour, we found that no jury was correct on all four of them. It appears that most jurors failed to absorb a great many of the judge's instructions and that the process of deliberation did not correct this problem.

Further evidence that the jurors learned less than they should have from the judge's instructions comes from examining the frequency with which different aspects of the law were discussed during deliberation. The instructions most often discussed involved points of law that the jurors were very likely to have heard about before they heard the case; thus, there is a strong possibility that much of their discussion

of the law was based not on the instructions they had heard from the judge but on prior knowledge.

For example, the element of the judge's instructions discussed most frequently was the definition of first-degree murder requiring premeditation and deliberation. Jurors were usually correct in their definitions (sixty-five correct statements, five incorrect, thirty-seven unclear). They very rarely went any further, however, in trying to define first-degree murder. There were only thirteen mentions of the definition of premeditation and deliberation, and only thirty-eight attempts to distinguish first-degree from second-degree murder (eight correct, four incorrect, twenty unclear, four questions, and two error corrections).

These results suggest that much of the jurors' discussion of the law on first-degree murder may have been based on the well-known phrase "premeditation and deliberation," and did not benefit from the new information provided by the judge's instructions. In addition, one might argue that the disproportionate amount of time spent discussing premeditation was inappropriate in this case, since fewer jurors favored first-degree murder than any other verdict choice.

Likewise, the familiar phrase "heat of passion" was the most commonly discussed element of manslaughter and accounted for 125 units, of which a third were incorrect or unclear statements. Interestingly, "involuntary manslaughter" was raised in ninety-three units, of which fifty-one were clearly incorrect. It is not surprising that most of the references were incorrect, since the judge had stated that the verdict category "involuntary manslaughter" was not relevant to this case. The fact that "involuntary manslaughter" was discussed almost as much as "heat of passion" in relation to the manslaughter verdict provides further evidence that juries rely at least as much on legal knowledge gained outside the courtroom as they do on the judge's instructions.

Although most of the law discussed by the jurors involved the substance of the verdict categories, jurors devoted 7 percent of their discussion of the law to the reasonable doubt standard, and 10 percent to the judge's instructions about



The fact that "involuntary manslaughter" was discussed almost as much as "heat of passion" in relation to the manslaughter verdict provides further evidence that juries rely at least as much on legal knowledge gained outside the courtroom as they do on the judge's instructions.

the jurors' duties. The juries' understanding of reasonable doubt and how they must rule in the face of reasonable doubt was extremely accurate. Not one person on any jury, however, raised the question of the definition of reasonable doubt.

Like "premeditated murder," the phrase "beyond a reasonable doubt" is one that is likely to be familiar to jurors from prior experience, so we cannot conclude that they learned this standard from the judge's instructions. Attempts to apply the reasonable doubt standard to the facts of the case were evenly divided between correct and incorrect/unclear applications. The reasonable doubt standard was almost always raised by jurors who were trying to persuade a harsher faction to move toward their position.

Procedural instructions were also used as arguing tactics. Of the 172 remarks made about jurors' duties, 114 were devoted to three of the eleven instructions given by the judge: that jurors should only be influenced by the evidence and law presented in court (forty-nine remarks); that jurors should not speculate about sustained objections (twenty-two); and that jurors should not consider the penalty or consequences of the verdict (forty-three).

These comments were also used primarily as a weapon to close off lines of argument that a juror disagreed with, and generally took the form, "We can't speculate about that," or "We're not allowed to consider that." Jurors applied these rules incorrectly thirty-nine times and were clearly incorrect forty-five times; only fifteen of these forty-five errors were corrected. A great deal of concern has been expressed about jurors' inability to disregard extra-evidentiary factors; our data suggest that this concern is appropriate. However, jurors may also use the judge's cautionary instructions to stifle discussion of unpalatable, but clearly relevant, evidence.

Conclusion

In summary, the process of deliberation seems to work quite well in bringing out the facts and arriving at a consensus about their sequence. Errors are corrected, and irrelevant facts and implausible scenarios are generally weeded out, at least in deliberations over this relatively simple homicide. The juries also do a good job of gradually narrowing down discussion to the important issues. On the whole, however, the discussion of the facts does not produce changes in votes, since jurors' verdict preferences in the case were rarely a function of a clear mistake on the facts.

Unfortunately, the jurors' understanding of the law was substantially inferior to their understanding of the facts and issues. The judge's instructions were not very effective in educating them in new areas, or even in focusing their attention on the meaning of the familiar terms. This failure to apply the law correctly was by no means a failure to take the law seriously. Discussions of the law took up one-fifth of the deliberation time and were carried out with great intensity, frequently with an apparent sense of frustration. The jurors understood that a key aspect of their task was to interpret the evidence in terms of the appropriate legal categories. They struggled to do so, but often failed.

There is no reason to believe that the jurors' misunderstanding of the law is a function of their mental capacities. It seems more plausible that the system is set up to promote misunderstanding. Factors blockading the serious jury trying to perform its task include: the convoluted, technical language; the dry and abstract presentation of the law following the vivid, concrete, and often lengthy presentation of evidence; the requirement that jurors interpret the evidence before they know what their verdict choices are; the fact that juries usually do not get copies of the instructions to take with them into the jury room; the lack of training in the law for jurors as part of their jury duty; the general failure to discover and correct jurors'

preconceptions about the law; the failure to inform jurors that they are allowed to ask for help with the instructions; and the fact that those who do ask for help are often disappointed by a simple repetition of the incomprehensible paragraph.

Research on jurors' comprehension of judge's instructions is increasing, but there is still very little. We do not even know whether juries that ask for help with the instructions do better than juries that try to muddle through on their own. Research on specific techniques for improving juror comprehension indicates that improvement is possible. At any rate, it seems profoundly unfair to criticize juries for failing to perform well a task that, by all the usual educational criteria, has been stacked against them.

LQN



Phoebe Ellsworth is the Kirkland and Ellis Professor of Law and Professor of Psychology at the University of Michigan. She currently is conducting research on emotions and interpretation of social situations across cultures. In addition to several Department of Psychology courses, she teaches Psychology of Litigation, Juries, and Social Science Research Methods at the Law School.