1988

An Experiment on the influence of structured instructions on jury decision making

Thomas Breitenbucher

The University of Montana

Let us know how access to this document benefits you.
Follow this and additional works at: https://scholarworks.umt.edu/etd

Recommended Citation
Breitenbucher, Thomas, 'An Experiment on the influence of structured instructions on jury decision making' (1988). Graduate Student Theses, Dissertations, & Professional Papers. 8655.
https://scholarworks.umt.edu/etd/8655

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mail.lib.umt.edu.
COPYRIGHT ACT OF 1976

This is an unpublished manuscript in which copyright subsists. Any further reprinting of its contents must be approved by the author.

Mansfield Library
University of Montana
Date: 1988
An Experiment on the Influence of Structured Instructions on Jury Decision Making

by

Thomas Breitenbuecher
B.A., University of Minnesota, 1973

Presented in partial fulfillment of the requirements for the degree of Master of Arts
University of Montana
1988

Approved by

Chairman, Board of Examiners

Dean, Graduate School

Date
An experiment was conducted to determine if requiring jurors to work with structured instructions making them decide a case one issue at a time, committing themselves to their partial decisions after each is made, would lead them to better apply the law. One hundred and fifty-six college "jurors" were involved. They were assigned to three groups, each group containing nine, six-person juries. Four juries were short one member, and one jury was short two members due to subjects failing to fulfill their experimental commitment. All groups heard an audio tape of a civil case involving product liability. One group heard a version of the case containing only the facts essential to decide the case. The other two groups heard a more lengthy presentation of the same case with legally irrelevant material added. The case presented was designed so that, if the legal instructions were followed, the verdict had to be for the defendant. Once they heard their version of the case, the juries in each group received different instructions. The group that received the stripped down version of the case received unstructured legal instructions. One of the groups that heard the lengthier case presentation also received unstructured instructions. The third group, also hearing the lengthier case presentation, received specific instructions telling them in what order they must decide the legal and factual issues involved. There were no statistically significant differences between the groups in the patterns of verdicts reached. There were also no significant differences in the behavior of the two groups hearing the longer case presentation, deliberating under structured or unstructured instructions, as measured by the confidence they felt in their verdicts nor in their evaluation of the evidence. Thus the hypothesis that structured instructions could aid in the deliberation of complex cases was unsupported. In the discussion of the experiment possible interpretations of negative results and questions for future research are considered.
ACKNOWLEDGEMENTS

Thanks to my committee, Drs. Doctor, Means, Silverman and Walters. They are what professional scholars should be but sometimes fall short of: learned and humane men.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION AND REVIEW OF THE LITERATURE</td>
<td>1</td>
</tr>
<tr>
<td>II. METHOD</td>
<td>24</td>
</tr>
<tr>
<td>III. RESULTS</td>
<td>31</td>
</tr>
<tr>
<td>IV. DISCUSSION</td>
<td>38</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>64</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>71</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION AND REVIEW OF THE LITERATURE

An experiment was conducted to determine if giving more structured instructions to jurors might improve the quality of their deliberations. In order to provide a context for this problem a brief review of relevant findings in the following areas is in order: 1) the history of legal psychology and jury research, 2) evidence concerning the ability of jurors to accomplish the task which the law requires, 3) research findings of cognitive psychology in the areas of problem solving, expert decision making, memory, confirmation bias, group polarization, and information integration theory, 4) a brief consideration of some problems of external validity in jury simulations.

HISTORY OF LEGAL RESEARCH

There is a long history of psychological research into legal issues, beginning with Muensterberg’s controversial book On the Witness Stand (1908). In that book he attempted to directly apply the principles of experimental psychology to the problems of law. To say that officers of the court were not impressed would be an understatement. The heart of Anglo American jurisprudence is precedent,
and any suggestion that things might be done differently because that would be more "scientific" has not been welcomed from the legal community. Robinson's book, Law and Lawyers (1935), which attempted by a slightly different route than Muensterberg, to reduce the problems of law to the problems and methods of psychology again contributed less light than heat. Yet Robinson's acknowledgement that it was not possible to simply apply existing laboratory findings to questions of law, was a step toward making a fruitful collaboration between the legal system and experimental psychology possible. Robinson acknowledged that psychology of law must be an applied psychology, not just an extension of general psychological laws to a new domain.

In the past lawyers and psychologists have tended to ignore one another (Fahr, 1961). The lawyers often believed that the laboratory research cited was of dubious application to the more complex milieu of the courtroom (Loh, 1981). The psychologists were puzzled at the indifference and hostility to their attempts to reform and improve the operation of the law. It is only since the beginning of the 1970s that experimental research in psychology and the law has become vigorous once again, following a long fallow period from the 1920s onward (Wells, 1984). Tapp's Law, Justice and the Individual
Society: Psychological and Legal Issues (1977), marks a turning point and “coming of age” in the relationship between lawyers and psychologists, marked by a greater sensitivity on the part of psychologists to the specific needs and point of view of those in the legal system (Loh, 1981). There is speculation, but no definite answer as to why experimental psychologists have become interested in the law once again (Wells & Loftus, 1984). The last ten years have seen a great increase in the growth of legal psychology with an emphasis on the part of psychologists on practical and applied work. At the same time lawyers “have grudgingly begun to believe that psychologists can make a practical contribution to the legal process” (Konecni and Ebeson, 1978, p. 39)

The Ability of Jurors to Accomplish Their Task

There has been a great deal of research on jury behavior since Muensterberg first pioneered the application of psychology to problems of law. A number of relatively recent reviews have attempted to consider and weigh the various strands of inquiry (Gerbasi, Zuckerman, and Reis, 1977; Saks and Hastie, 1978; Nemeth, 1981; Monahan and Loftus, 1982) One of the most active areas of inquiry has been an examination of factors affecting jury deliberations, and whether in fact it can be presumed that the average juror is capable of performing the task which
the law expects. A juror takes an oath to decide a case upon the law and the evidence presented in court. His personal opinions or sense of what the law should be are considered legally irrelevant. While this may be what the law demands, many researchers have been led to ask what empirical foundations there may be for the belief that this is possible, let alone usual. Kalven and Zeisel found that judges and juries disagreed on verdicts in approximately 20% of cases (1966). There are two major areas of concern here. One is the question of how prejudice and personality traits might affect trial outcomes. The second concerns the question of how well jurors apply the law. As we shall see the two questions in practice can't be very easily separated. One question is almost the mirror image of the other. When jurors are not applying the law they may be assumed to be applying their biases.

A major function of the judge is to interpret the law. The task of the jury is to determine the facts and apply the law. Yet there is a great deal of evidence that the jury does not always follow the law as set down by the judge. Sometimes they don't understand the law. Hervey (1947) found after examining questionnaires returned by 185 jurors that 40% had not understood the judge's instructions. Strawn and Buchanan (1976) found that only
50% of simulated jurors understood that the defendant did not have to prove that he was innocent. Overall, subjects in the Strawn and Buchanan study showed an average of only 70% comprehension when tested on their knowledge of standard legal instructions, immediately after being given them. A study by Elwork, Sales and Alfini (1977) found that juror comprehension and memory for the standard legal instructions were as effective in conveying legal concepts as receiving no instructions at all. Simon (1967) in her study of the insanity defense found that of twelve mock juries asked to consider a case using the M’Naghten rule, only eight formally considered the defendant’s ability to tell right from wrong, though that is a crucial part of the rule. McGlynn and Dreiling (1981) found evidence that a jury will tend to consider a defendant less insane the more serious the crime, when an insanity plea is entered. However, this is directly contrary to the demands of the law.

Sometimes jurors seem to prefer their own version of the law. Kalven and Zeisel (1966) concluded that the modification of the law to conform to what the jury believed it ought to be, was a factor in 50% of the disagreements about verdicts between judge and jury. The authors go on to outline such matters as self-defense, contributory fault and liability damages as areas where
juries often follow their own version of the law

Monahan and Loftus (1982) cite a number of factors which have been investigated as possible sources of bias in jury deliberations: these include defendant characteristics, juror personality, the joining of issues in a single trial and procedural rules.

A question of great current interest is whether a jury can be expected to understand the issues, and apply the laws which are part of contemporary complex civil litigation (Austin, 1984). As law becomes more complex, in reflection of the increasing complexity of society itself, it may well be asked whether it is reasonable to expect the average person to be able to comprehend and deal with such complexity when called upon for jury duty. Former Chief Justice Burger (1982) has written, "It borders on cruelty to draft people to sit for long periods trying to cope with issues largely beyond their grasp" (p. 21). As Kalven (1964) states, many doubts have been expressed about the possibility of reliable fact-finding by, "12 persons brought in from the street, selected in various ways, for their lack of general ability..." (p. 1068). The Third Circuit Court ruled in 1980 that, "due process precludes trial by jury when a jury is unable to perform [its] task with a reasonable understanding of the evidence and legal rules." (Austin, 1984, p.1) The Ninth Circuit

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Court, however, refused to accept the possibility of there ever being a case "so overwhelmingly complex that it is beyond the abilities of a jury." (Austin, 1984, p.1)

Despite the obvious practical and constitutional importance of this issue, and although there has been a great deal of research in various aspects of jury behavior, there has been a paucity of empirical research on jury comprehension. There has not only been relatively little research into jury comprehension, there has been even less research into ways of improving the ability of a jury to apply the instructions received from the bench. (Severance & Loftus, 1982).

The focus of this experiment was to determine if more structured instructions from the bench might lead to more accurate application of the law. The idea for the experiment came from the Sharon libel case against Time magazine. In that case the judge directed the jury to consider the three legal elements necessary for libel one at a time. The jury had to decide: was the Time story untrue, was it damaging to the reputation of Sharon, and was Time either malicious or careless in publishing it. The jury had to announce its decisions publicly after each was made. Considering the fact that a number of jurors in their public statements after the trial expressed distaste for Time magazine’s conduct, the ultimate verdict of
acquittal is remarkable though in good conformance with the law. In the February 4th issue of *Time*, Floyd Abrams, a New York lawyer who specializes in First Amendment cases is quoted as saying of the case, "Here the jury distinguished between a mistake and a lie. Most juries have not been able to do that, and consequently we have a 75% reversal rate of verdicts for the plaintiff on appeal" (p. 66).

The outcome of the Sharon trial brought to mind three related elements in the research literature. First, the judge's instructions to the jury can be a crucial element in the outcome of a trial. Oros and Elman demonstrated the crucial effect the judge's instructions to the jury can have in rape trials (1979).

Second, there is also some evidence that jurors are most likely to be guided by their personal biases and opinions when the evidence is ambiguous (Kaplan and Miller, 1978). In another study Kaplan and Miller state, "our research on trait biases suggest that when evidence is ambiguous or of questionable reliability, biases remain operative" (1974, p.52). An analogous result was reported by Sue, Smith and Caldwell who found that when the remaining evidence is weak, jurors will disregard the judge's instructions and be influenced by inadmissible evidence (1973). It seems as if in a state of uncertainty the jury will use whatever
information comes to hand. In Simon's book on the insanity plea she records long wandering discussions in the jury room as one juror after another related personal experiences, stories and recollections as they attempted to come to grips with what was meant by the difficult and somewhat obscure concept of legal insanity. While this is understandable, it is not what the law envisions jury deliberations should be. It seems reasonable that any procedures that can be invented to make the structure of evidence and law clearer, would aid in deliberations.

There is a third element of the research literature which the Sharon case brings to attention. There is evidence that jurors have trouble separating issues and evidence when there are several issues at stake in a trial. When several criminal offenses are joined in a trial, experimentation has shown that ratings of the defendant's guilt are higher than when offenses are tried separately (Horowitz, Borden and Feldman, 1980; Tanford and Penrod, 1982). There is support for the hypothesis that one of the reasons for this phenomenon is that evidence relevant to one charge may be connected with another offense (Tanford and Penrod, 1982; Tanford, 1985). It also appears that there is a tendency to remember only evidence relevant to the verdict preferred by a juror (Greene, 1981).
The apparent success of the judge's instructions in the Sharon case, in aiding the jury in their complex deliberations, leads to the question of whether similar instructions might be helpful in reducing ambiguity, and aiding the separation of issues in other types of civil cases. There is considerable ambiguity at the end of any trial. A great deal of evidence will have been presented that must be integrated by the jury. One of the most important tasks of the jury is to extract relevant information from very complex and confused context (Ebeson & Konecni, 1980, Gerbasi, Zuckerman & Reis, 1977). It seems possible that a set of instructions that would focus the case on specific issues, demanding that they be decided one at a time, would serve to reduce the amount of ambiguity perceived by the jury and thus the amount of bias in applying the law.

Relevant Findings of Cognitive Psychology

Some of the findings of cognitive psychology, though not directly applicable to the problem set out here, offer a good deal that is suggestive. Much research has been devoted to the study of how people solve well defined problems with clearly specified goals. The jury's task is not a well defined problem (Reitman, 1964). Unlike a chess problem, neither the initial state of affairs (the validity of the evidence) nor the procedures by which the
verdict must be reached are precisely specified. The goal is to determine whether there is enough evidence to meet the legal criteria offered in the judicial instructions. Ill defined problems are generally more difficult to solve than well defined ones (Chi and Glasser, 1985).

Newell and Simon's (1972) classic work tries to set out the universal elements of problem solution. First one must recognize the task environment: the terms in which the problem is presented. This may consists of words, symbols, meanings or objects. Second, one must transform that information into a problem space: a mental representation of the task. The problem space is one's interpretation of the initial state, the goal and the acceptable means to attain the end. Once the problem space is defined one relies on information in memory to assess whether or not various possible steps can bring one closer to the desired goal. One can conceive of the initial state of the jury's problem as being the evidence and the law to be applied. The jury must sort through the evidence, finding the information relevant to each part of the law, weigh that evidence, and come to a decision as to whether the criteria of the law have been met. At its simplest, the jury's task is rather like a logic problem where the task is to sort through a great deal of relevant and irrelevant information to determine if there is sufficient relevant
information to make a certain inference. The juror's prejudices and intuitions might be considered simply another form of irrelevant information which must be sorted through.

In a real court case the juror's task is complicated by the fact that the truthfulness of the evidence presented can not be taken for granted. The juror must make an assessment of the probability of the testimony he hears.

The question of how people assess probabilities is a complex one, though the errors that are made are often of systematic and predictable types (Lichtenstein, Slovic, Fischoff, Layman & Combs, 1978). For the sake of simplicity, the present study will focus on the question of how the cognitive process of sorting relevant information and applying it might be enhanced. It seems plausible that an assessment of probabilities must rest on the ability to extract and apply relevant information. Information gained about cognitive processes might be relevant to future research into the larger question of how jurors establish the credibility of evidence.

The function of structured jury instructions might be seen as that of offering to the juror the problem solving heuristic which Newell and Simon (1972) suggest, of creating subgoals. The structured instructions divide the jury's work into a series of subtasks, which are more
easily managed.

In a series of studies, Voss and his colleagues attempted to discover how people go about solving ill defined problems (Voss, Greene, Post, & Penner, 1983; Voss, Tyler and Yengo, 1983). Their work seems to indicate that people usually try to solve ill defined problems by breaking them down into smaller more manageable subproblems. This process is precisely what the structured jury instructions considered in this study attempted to aid.

The analogy between the jury process and conventional problem solving tasks is not entirely satisfactory, however. Since neither the initial state, the permissible moves to the goal, nor the goal itself are precisely defined, the analogy is somewhat strained. A better paradigm might be found in considering the work that has been done in examining the formation and application of expert knowledge. The juror's task is rather like that of a doctor, who examines a patient presenting a variety of symptoms. The physician must decide which of the many things he observes and is told are relevant. Some physicians are far better at diagnosis than others. What is the difference between an expert, a beginner or a bungler? According to Larkin and Reif (1979) the answer seems to be that the knowledge of an expert is
hierarchically organized. That is, expert knowledge logically organizes details and phenomena into logically grouped chunks.

Paul Johnson's study of expert knowledge has found that among other things, the expert has a "high altitude overview." He goes on to say, "Highly expert people have an enormously efficient picture of what they are trying to do- they see things in perspective so they know just what additional information they need. The less expert individual collects a great deal of unnecessary data, because he doesn't know what he needs to know." (Hunt, 1982, p.264)

Chi, Glasser and Rees (1982), in a study of how experts and novices in physics classify problems found that novices were more influenced by the appearance of a problem than the physical principles involved. This would seem to indicate that one of the things experts are able to do, which makes their judgments more accurate, is see beneath the superficial features of a problem. It seems plausible that what structured instructions do is help organize information into chunks, helping the jurors get at the underlying principles of the case. When judicial instructions are unstructured, they are simply additional pieces of information to be sorted out. They may not necessarily aid in organizing the legal problem into the
most efficient and accurate structure.

It is relevant here to recall research done on memory retrieval. Bartlett (1932) contended that storage and retrieval of verbal information is "effort after meaning." That is, people do not remember, after a relatively brief interval, exactly what they have read or heard, but what they perceive to be its meaning. These findings were confirmed by several researchers testing the ability of subjects to exactly recall sentences (Sachs, 1967; McKoon, 1977; Anderson and Paulson, 1977). What the subject recalls is not the literal order of words, but the meaning of a sentence.

But the fact that we recall what we believe is meaningful implies that our present beliefs about the material to be recalled can affect what we remember. Snyder and Uranowitz (1978) and Halpern (1984) have conducted research on the effect of stereotypes on memory, finding that, as Halpern expressed it, "It seems that it is our memories that change while the stereotype resists change" (p.252). Likewise, the legal system has come to recognize that the testimony of eyewitnesses is very often tainted by what the witness has learned after the event (Loftus, 1975; Loftus, Miller & Burns, 1978). The implication would seem to be that how one conceives of a case in the present can affect what is remembered about
it. Changing the structure of jury instructions might well affect the conception of the case as it is being deliberated, and this could certainly affect what is remembered. What is remembered must certainly have an affect on the outcome of deliberations.

A common fallacy of thought, even where all information necessary to make a decision is present and readily available is to fail to consider, and thus ignore, disconfirming information. Studies by Smedlund (1963) and Wason (1960, 1968) have demonstrated that confirmation bias is a prevalent fallacy of everyday reasoning. The phenomenon has even been demonstrated in NASA scientists (Mynatt, Doherty, & Tweney, 1978). Yet work by Janis and Mann (1977) suggests that poor decisions are often the result of failure to think through all relevant considerations.

The research on confirmation bias would lead one to predict that members of a jury will seek to marshal evidence that confirms their initial impression of the case rather than seeking systematically to determine whether their impression is warranted by the evidence. There is evidence that the jury's eventual decision often is in the direction of the initial vote and that discussion tends to make those initially in favor of a position, more strongly in favor after deliberation.
(Kalven & Zeisel, 1966; Davis, 1973; Meyers and Kaplan, 1976; Kaplan and Schersching, 1981). This is a special case of the phenomenon of group polarization (Cartwright, 1973; Lamm & Myers 1978).

There is a great deal of evidence that the tendency of a jury to move in the direction of the majority opinion is largely affected by the majority producing proportionately more evidence and argument to support their opinion (Kaplan, 1984; Kaplan & Miller, 1983; Lamm & Myers, 1978, Stasser, Stella, Hanna, & Colella, 1984). It might be theorized that the effect of this accumulation of supportive evidence is to produce an impression of a thorough examination of the issues and overwhelming agreement of the case evidence. The effect of confirmation bias thus is cumulative and pushes the jury toward unanimity in the direction of the majority view. Structured instructions, by forcing consideration of each of the relevant legal issues, one at a time, works against confirmation bias. The procedure is in the spirit of the recommendation of Stasser et al. (1984) that jury deliberation be structured to "procedurally compensate for the advantage of the majority by encouraging an unbiased review of evidence" (p. 125). Structured instructions assure that all of the issues are systematically deliberated. Confirmation bias would lead one to expect
that in unstructured deliberations only those issues that are favorable to the majority opinion will be considered.

Information integration theory postulates that evidence has both scale and weight (Anderson, 1976; Kaplan & Miller, 1979). Scale refers to the importance of the evidence in terms of the judgment to be made. If the defendant is an expert marksman, that would have high scale value if he is charged with shooting his rich uncle. If he was in another county when the shooting occurred, the scale value would be low. Weight refers to the importance of the information in making the particular judgment. One would hope that the fact that the defendant has an honest face would have less weight than, for instance, the fact that he was at the scene of the crime.

In a typical court case, the content of testimony provides scale values, while its credibility and relevance to the verdict affect its weight. Each piece of information is integrated into the final decision in a kind of weighted average. As Kaplan and Miller note, "the most interesting property of the weighted average form is that the weighting is relative: as the weighting of one element decreases...the weight of the other elements will increase" (p. 56).

One might postulate that one of the things that might happen in giving a jury structured instructions is that
the weight of various elements of the evidence would shift. In focusing jury discussion down to several specific issues to be decided in sequence, the jury would be led to see the relevance of the evidence to the law more clearly. Thus the evidence would have more weight. And there would be less need for the members to draw on folklore and personal experience to resolve the ambiguities of the case.

External Validity of Individual Measures in Jury Research

The question of whether one should test hypotheses concerning jury behavior by measuring the responses of individuals, or whether only the collective responses of groups should be used, is controversial. Erlanger has argued there seems to be little need or justification at this point for simulated studies using individuals to decide cases (1970). The use of individuals rather than groups in simulations has come under a fair amount of criticism (Konecni and Ebbsen, 1980; Bermant, McGuire, McKinley, & Salo, 1974). The methods which Strodtbeck used in the 50s, gaining cooperation with the courts, using members from the jury pool, who acted as a jury in an actual court setting, seem ideal. (Strodtbeck, James, & Hawkins, 1957).

However, Bray and Kerr (1979) note, "The difficulty and cost of carrying out highly realistic simulations is
simply prohibitive for many investigators" (p. 108). And thus in their review of current research they found that 70% of the 72 studies reviewed, took place in the laboratory. In the instance of the present study it would be nearly impossible to convene over 150 juries of 6 persons each in realistic trial settings in order to determine if there is a significant difference in trial outcomes through the use of structured instructions.

This controversy represents a considerable dilemma for the jury researcher. On the one hand it seems compelling, since the jury represents a group process, to use experimental procedures which involve presenting cases to groups and having them deliberate to a result. The measure used then would be the difference in number of verdicts pro and con produced between one experimental condition and another. But if the unit of observation is the group, and the measure is "guilty or not", then one must essentially multiply the number of subjects needed by the number of members in the jury. If 150 observational units are needed to test a hypothesis, and there are six persons to a jury, then 900 subjects are needed.

The alternative procedure used in jury research is to present case materials to individuals and have them come to a conclusion without deliberation. This procedure is efficient in time and numbers of subjects required, but it
obviously ignores the element of group process in jury deliberations.

The method used in this study represents a compromise which implements procedures from both approaches. Case materials were presented to mock jurors and they deliberated for a period of time under no requirement to come to a unanimous verdict. At the end of that time individual jurors were polled for a verdict and the jurors were surveyed for their reaction to the case materials. The statistical analysis of results maintained the group as the unit of analysis. That is, individual yea and nea votes were not simply pooled and compared between conditions, but an analysis of variance was conducted using each mock jury as the unit of observation. Individual verdicts were expressed as a fraction of total voting for the plaintiff over total jury members, and rated evaluations of the importance of evidence were averaged for each jury and those averages were compared by one way analysis of variance across the experimental conditions.

It can be reasonably argued that in areas of inquiry where there has been relatively little research, such as the effect of judge's instructions, there still is a point in doing research where single individuals, after a period of deliberation, are asked to render a verdict. Such
studies can give some idea of how useful it might be to attempt to further study under more realistic conditions. If individuals deliberating with structured instructions come to significantly different conclusions than individuals who do not, then it would be plausible to look for similar differences in groups operating under conditions more like those of an actual trial. The finding of no significance does not rule out a possible significant effect with actual juries, nor does the finding of differences in individuals guarantee that such differences will be found in groups. Nevertheless, the extremely large numbers involved in working with mock juries, where each group counts as a single subject, might argue for initially examining individual responses to try to get a rough measure of the magnitude of the effect to be expected.

There is empirical support for this procedure in the findings of Kalven and Zeisel (1966) that 9 out of 10 juries reached the verdict favored by the majority on the first ballot. Several mock trial studies have found that proportionate strength of the majority position was a good predictor of the eventual outcome (Davis, 1973; Davis, Kerr, Stasser, Meek, & Holt, 1977; Kerr, Nerenz, Herrick, 1979).
Summary of Experimental Hypothesis

The purpose of the study then, was to examine the question of whether or not structured instructions would aid jurors in accurately applying the law to complex cases. If a group of jurors deliberating with structured instructions is, on average, more successful in applying the law to a complex case and in separating out relevant from irrelevant evidence than jurors working with unstructured instructions, this would provide support for the hypothesis that structured instructions contribute to accuracy of the jury process.

If the groups receiving structured instructions perform in a way that is not statistically different from groups receiving unstructured instructions, then the hypothesis that structured instructions are useful in helping jurors accurately perform their duties is unsupported.
CHAPTER II

METHOD

Subjects

A total of 156 subjects, 76 males and 80 females, drawn from the introductory psychology course at the University of Montana, participated in the experiment. They participated as part of a course requirement. No mock jury was composed exclusively of males or females. Subjects were randomly assigned to one of three experimental cells. In the first, 52 subjects (25 males and 27 females in seven groups of 6 and two groups of 5) comprised the condition hearing a shortened version of a product liability case. This will be referred to henceforth as the short cell. In the second experimental cell, 53 subjects (26 males and 27 females in eight groups of 6 and one of 5) participated in the condition which deliberated under unstructured instructions. This will be referred to as the unstructured cell. In the third, 51 subjects deliberated under structured instructions. This third set was composed of 25 males and 26 females divided among seven groups of 6, one of 5 and one group of 4. These groups will henceforth be referred to as the structured cell. The reason why 5 juries were short a member and one was short two members was that some subjects did not fulfill their commitment to participate.
in the experiment. As the unit of observation was the group, the shortfall of these members was not judged to be a serious problem.

**Apparatus**

An audio taped mock trial was used, so constructed that a verdict for the defendant was logically required if the legal principles given in the judicial instructions were accurately applied to the case. The case involved product liability (See Appendix A for the trial transcript). The trial tape came in two versions. One version contained only the logically essential elements of the case (see Appendix B). In this version there was only a single argument and counterargument for each of the three major issues to be decided. Further, the plaintiff’s injuries and reverses, were covered in two short statements rather than being presented at length. A lengthier version of the case (see Appendix A) presented to juries in two experimental cells forced the jurors to sort through a great deal of information, much of which was not relevant to the legal issues to be decided, in order to come to a verdict.

**Procedure**

One hundred and sixty two subjects heard an audio taped civil trial involving the product liability case described above (See Appendices A and B). There were three
experimental cells each consisting of 9 juries (1 jury had four members, 5 had five members and 23 had six members). Each jury heard the audio taped case material and discussed it for 20 minutes. At the end of that time each juror voted on the verdict to be handed down. There were two different sets of instructions directing the jurors in how they were to proceed with their deliberations.

The arrangement of case types and instruction types among the three groups can be diagrammed as follows:

I. SHORT II. UNSTRUCTURED III. STRUCTURED

<table>
<thead>
<tr>
<th>Bare Bones Case-</th>
<th>Full Case-Structured Instructions</th>
<th>Full Case-Unstructured Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstructured Instructions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The short cell heard the stripped down version of the case. For this group the task of sifting through the conflicting claims of the case, to get to the core of information necessary to decide the legal issue was done for them. The purpose of this group was to provide a baseline against which the behavior of the other two groups could be judged. It was predicted that the pattern of verdicts and the way in which evidence was evaluated would be most similar between this group and the structured condition, thus providing further evidence that the structured instructions aided the jurors in getting to
the heart of the legal issues. It was expected that providing structure would, in effect, act in much the same way as decreasing the quantity of information that must be considered.

The structured cell heard an expanded version of the case (see Appendix A). In deliberating the case they received structured instructions, telling them in what order to decide the legal issues one by one. After they made each decision they were required to mark a jury ballot before they could go on to decide the next issue. It was hoped that the process of formally narrowing the discussion of issues, by means of structured instructions would aid the jurors in finding and using the relevant information in the case materials.

The third group heard the same case materials as the second group. They, however, were not told in what order to decide the issues involved during their deliberations. They were simply told the law, the legal issues to be determined and left to their own devices as to how to apply the law to the facts. This is, in fact, closely similar to typical courtroom procedure.

**MEASURES**

There were three types of measures used to examine differences in group behavior. The measures were: 1) differences in verdicts arrived at, 2) confidence in the
verdict and 3) evaluation of the evidence. The first measure was of the differences in verdicts among the three experimental conditions. The verdict arrived at in each group was expressed as a fraction i.e., the total voting for the plaintiff over the total number of individuals in the group. The reason why the verdicts were expressed in this way rather than pooling together the total number of votes for the plaintiff and comparing those totals across the three experimental conditions, was that the unit of measurement in this experiment was the group rather than the individual. A jury is not simply an aggregate of individuals. The juror works in a group to come to a decision and the group process will greatly influence him or her. The individual's predilections, prejudices and his willingness to act on them will be modified to some degree by the others in the group. Thus it seemed inappropriate to statistically examine the data from this experiment as if one were dealing with the behavior of 156 unconnected individuals. Each one of those individuals was part of a particular group and their behavior may have been different if they had been attached to some other group. The statistical procedures used must necessarily reflect that fact. The experiment was concerned with examining the behavior of 27 groups rather than 156 individuals.
In that not all groups had equal numbers of subjects, the best way to reflect that fact was to express the behavior of each group as a fraction which divided those voting for the plaintiff over the total number in the group. The alternative of simply using the total votes for the plaintiff in each jury as the measure of their result would lose the information about unequal group sizes. Five votes for a plaintiff in a group of five means something different than five votes out of a group of six.

The second measure, the confidence that jury members had in their verdicts was assessed on a five point Likert scale where response possibilities ranged from "extremely confident" to "not very confident" (see Appendix F).

The third measure in the experiment, that of differences in evaluation of the evidence, was done by having each juror fill out a questionnaire which listed each piece of evidence and asked them to evaluate on a 5 point Likert scale how important each was in arriving at their verdict (see appendices D and E). The responses on the questionnaires were then analyzed in the following fashion: first, evidence was divided into two major categories: that which was legally relevant and that which was legally irrelevant. The irrelevant evidence consisted of information about the seriousness of the plaintiff's injuries and his financial reverses. This evidence is
defined as "irrelevant" because, as the case was constructed, it had no bearing on the legal issues to be decided. That the plaintiff was injured was not contested. That he suffered was not in dispute. What was at issue, was who was responsible for the injuries and why. Information about the plaintiff’s suffering was designed to sway the jury’s emotions, but it had no bearing on the technical legal issues.

Legally "relevant" information was so defined because it had bearing on the three legal issues that had to be decided. The category of relevant evidence was divided into three major categories, each having two subcategories. The first category (A) comprised evidence germane to the issue of legal responsibility. Its subcategories were (1) evidence supporting legal responsibility and (2) evidence diminishing legal responsibility. The second category (B), comprised evidence germane to the issues of whether there was a defect in the product. Its subcategories were (1) evidence supporting the existence of a defect and, (2) evidence against the existence of a defect. The third category (C), comprised evidence germane to the issue of whether the defect caused the accident. Its subcategories were (1) evidence supporting the claim that a defect caused the accident and (2) evidence counter to that claim.
CHAPTER III

RESULTS

Given that the major thrust of the experiment was the comparison of the effect of structured instructions on jury deliberation an initial comparison was conducted between the structured and unstructured groups on each measure to estimate the probability that the two cells had a different population mean. Whatever other patterns there might be in the data, the absence of significant differences between these two groups render that data equivocal and difficult to interpret. That comparison thus must be the foundation of analysis. As the F and t tests are equivalent when there is only one degree of freedom, the F statistic was used as a matter of convenience in analyzing the data. There was no significant difference at the .10 level, between the verdicts for the plaintiff in the structured (M=.16) and unstructured group (M=.05), F (1,16)=.82. The Likert scale measures of the confidence in the verdict between the structured group (M=4.25) and the unstructured group (M=4.43) was not significant at the .10 level, F (1,16)= .46.

The Likert scale survey measures of the importance jurors attached to each piece of evidence also revealed no
significant difference at the .10 level. Table 1 contains a summary of these results. It presents a statistical comparison of the mean rated evaluation of three categories and twelve subcategories of evidence between the two experimental cells. The evidence, it will be recalled, was rated on a 5 point Likert scale in terms of its importance for arriving at the verdict which the group rendered. On that scale a rating of 1 meant the evidence was "not important," a rating of 5 meant it was "extremely important" (see appendices D and E). In the table standard deviations are recorded in parenthesis next to their respective means.

Insert Table 1 about here

Since the above results were negative without exception, it was decided to engage in a pairwise comparison of all groups on all measures: verdicts reached, confidence in verdicts and evaluation of evidence. The purpose of doing this was to determine if such comparisons might suggest reasons for the lack of significant results.

The statistical procedure chosen for these multiple comparisons was the Fisher significant difference test. While it provides protection against Type I errors, it is
TABLE 1

<table>
<thead>
<tr>
<th></th>
<th>Structured Mean (SD)</th>
<th>Unstructured Mean (Pooled SD)</th>
<th>F (1,16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ir</td>
<td>2.58 (.40)</td>
<td>2.64 (.59)</td>
<td>.05</td>
</tr>
<tr>
<td>R</td>
<td>3.46 (.26)</td>
<td>3.47 (.44)</td>
<td>.01</td>
</tr>
<tr>
<td>Ir/R</td>
<td>.76 (.10)</td>
<td>.78 (.14)</td>
<td>.05</td>
</tr>
<tr>
<td>D</td>
<td>3.11 (.30)</td>
<td>2.85 (.47)</td>
<td>1.92</td>
</tr>
<tr>
<td>D-</td>
<td>3.03 (.35)</td>
<td>2.83 (.59)</td>
<td>.74</td>
</tr>
<tr>
<td>D+</td>
<td>3.20 (.58)</td>
<td>2.88 (.48)</td>
<td>1.63</td>
</tr>
<tr>
<td>C</td>
<td>3.52 (.35)</td>
<td>3.69 (.29)</td>
<td>1.28</td>
</tr>
<tr>
<td>C-</td>
<td>3.73 (.18)</td>
<td>3.88 (.24)</td>
<td>2.29</td>
</tr>
<tr>
<td>C+</td>
<td>3.31 (.64)</td>
<td>3.50 (.39)</td>
<td>.60</td>
</tr>
<tr>
<td>L</td>
<td>3.69 (.49)</td>
<td>3.72 (.51)</td>
<td>.02</td>
</tr>
<tr>
<td>L-</td>
<td>3.96 (.60)</td>
<td>4.05 (.46)</td>
<td>.13</td>
</tr>
<tr>
<td>L+</td>
<td>3.41 (.62)</td>
<td>3.38 (.70)</td>
<td>.01</td>
</tr>
</tbody>
</table>

Ir = average of the scores measuring the importance of irrelevant evidence (evidence concerning the seriousness of Mr. Mars' injuries and financial losses).
R = average of the evaluations of the importance of all other case evidence.
Ir/R = the ratio between the two above measures.

D = all evidence regarding whether there was a defect in the plane. D = (D- + D+) /2.
D- = all evidence against the existence of a defect.
D+ = all evidence supporting the existence of a defect.

C = all evidence germane to the question of whether a possible defect caused the accident. C = (C- + C+) /2
C- = all evidence against the proposition that a defect in the plane caused the accident.
C+ = all evidence supporting the claim that a defect in the plane caused the accident.

L = all evidence concerning whether the defendant has any legal responsibility to the plaintiff, given that he was not the purchaser of the aircraft. L = (L- + L+) /2
L- = all evidence against the existence of a legal obligation of the manufacturer.
L+ = all evidence supporting the existence of a legal liability.
sufficiently liberal to be appropriate for an area of

* indicates significance at the .05 level
study in an early stage of research. A statistical test that was too stringent might not allow suggestive and tentative patterns in the data to emerge.

The initial analysis of variance among the three group verdict means, $F(2,24) = .69$, showed no difference at the .05 level. Thus, according to the procedure for the Fisher’s test, no further comparisons were made.

There were no Likert scale measures of confidence in the verdict for the short groups as that measure was not a part of the original experimental design and was added after the short groups had been run.

The pairwise comparisons of how the three experimental cells evaluated the evidence is best presented in four tables. In table 2, the results of the one way analysis of variance among all three group means is recorded for all measures of evidence evaluation. In tables 3 and 4, and 5 comparisons among the means, and the Fisher’s significant difference statistic are listed for those measures which reached significance in the initial analysis of variance. That is, if the comparison among the three means on a particular measure did not reach significance at the .05 level, pairwise comparisons between groups on that measure were not undertaken (although for the sake of completeness the results of F test comparisons of all measures is contained in
In that comparison by jury groups failed to find significant differences, it was questioned whether comparison of verdicts using individuals as the unit of measurement might show a trend that would be suggestive. An analysis of variance among the three groups showed a difference significant at the .05 level, $F (2,155) = 4.51$. Pairwise comparisons among the three groups showed no significant difference between the structured and unstructured groups nor between the short and unstructured group and a significant difference at the .05 level between the short group and the structured group $FSD (51 \ df) = .154$.
### TABLE 2

**Comparison of Likert Ratings of Evidence Importance Between the Unstructured, Short and Structured Groups**

<table>
<thead>
<tr>
<th></th>
<th>Unstructured Mean (SD)</th>
<th>Short Mean (SD)</th>
<th>Structured Mean (SD)</th>
<th>F(2,24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ir</td>
<td>2.64 (.59)</td>
<td>2.49 (.59)</td>
<td>2.58 (.40)</td>
<td>.17</td>
</tr>
<tr>
<td>R</td>
<td>3.47 (.44)</td>
<td>3.41 (.30)</td>
<td>3.46 (.26)</td>
<td>.08</td>
</tr>
<tr>
<td>Ir/R</td>
<td>.78 (.14)</td>
<td>.73 (.15)</td>
<td>.76 (.10)</td>
<td>.12</td>
</tr>
<tr>
<td>D</td>
<td>2.85 (.47)</td>
<td>2.80 (.35)</td>
<td>3.11 (.30)</td>
<td>1.70</td>
</tr>
<tr>
<td>D-</td>
<td>2.83 (.59)</td>
<td>3.05 (.20)</td>
<td>3.03 (.35)</td>
<td>.77</td>
</tr>
<tr>
<td>D+</td>
<td>2.88 (.48)</td>
<td>2.55 (.61)</td>
<td>3.20 (.58)</td>
<td>2.98</td>
</tr>
<tr>
<td>C</td>
<td>3.69 (.29)</td>
<td>3.86 (.20)</td>
<td>3.52 (.35)</td>
<td>3.31</td>
</tr>
<tr>
<td>C-</td>
<td>3.88 (.24)</td>
<td>4.67 (.21)</td>
<td>3.73 (.18)</td>
<td>51.62*</td>
</tr>
<tr>
<td>C+</td>
<td>3.50 (.39)</td>
<td>3.06 (.57)</td>
<td>3.31 (.64)</td>
<td>1.50</td>
</tr>
<tr>
<td>L</td>
<td>3.72 (.51)</td>
<td>3.04 (.46)</td>
<td>3.69 (.61)</td>
<td>4.23*</td>
</tr>
<tr>
<td>L-</td>
<td>4.05 (.46)</td>
<td>2.98 (.60)</td>
<td>3.96 (.60)</td>
<td>10.17*</td>
</tr>
<tr>
<td>L+</td>
<td>3.38 (.70)</td>
<td>3.09 (.90)</td>
<td>3.41 (.62)</td>
<td>.52</td>
</tr>
</tbody>
</table>

Ir = average of the scores measuring the importance of irrelevant evidence (evidence concerning the seriousness of Mr. Mars' injuries and financial losses).

R = average of the evaluations of the importance of all other case evidence.

Ir/R = the ratio between the two above measures.

D = all evidence regarding whether there was a defect in the plane. D = (D- + D+)/2.

D- = all evidence against the existence of a defect.

D+ = all evidence supporting the existence of a defect.

C = all evidence germane to the question of whether a possible defect caused the accident. C = (C- + C+)/2

C- = all evidence against the proposition that a defect in the plane caused the accident.

C+ = all evidence supporting the claim that a defect in the plane caused the accident.

L = all evidence concerning whether the defendant has any legal responsibility to the plaintiff, given that he was not the purchaser of the aircraft. L = (L- + L+)/2

L- = all evidence against the existence of a legal obligation of the manufacturer.

L+ = all evidence supporting the existence of a legal liability.
* indicates significance at the .05 level

**TABLE 3**

Comparison of Likert Ratings of Evidence Importance Between the Structured Group and Short Group

<table>
<thead>
<tr>
<th></th>
<th>Structured Mean (SD)</th>
<th>Short Mean (SD)</th>
<th>FSD, (24 df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-</td>
<td>3.73 (.18)</td>
<td>4.67 (.21)</td>
<td>.94*</td>
</tr>
<tr>
<td>L</td>
<td>3.69 (.61)</td>
<td>3.04 (.72)</td>
<td>.65*</td>
</tr>
<tr>
<td>L-</td>
<td>3.96 (.60)</td>
<td>2.97 (.60)</td>
<td>.99*</td>
</tr>
</tbody>
</table>

**TABLE 4**

Comparison of Likert Ratings of Evidence Importance Between the Unstructured Group and Short Group

<table>
<thead>
<tr>
<th></th>
<th>Unstructured Mean (SD)</th>
<th>Short Mean (SD)</th>
<th>FSD (24 df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-</td>
<td>3.88 (.24)</td>
<td>4.67 (.21)</td>
<td>.79*</td>
</tr>
<tr>
<td>L</td>
<td>3.72 (.51)</td>
<td>3.04 (.46)</td>
<td>.68*</td>
</tr>
<tr>
<td>L-</td>
<td>4.05 (.46)</td>
<td>2.98 (.60)</td>
<td>1.07*</td>
</tr>
</tbody>
</table>

**TABLE 5**

Comparison of Likert Ratings of Evidence Importance Between the Structured Group and Unstructured Group

<table>
<thead>
<tr>
<th></th>
<th>Structured Mean (SD)</th>
<th>Unstructured Mean (SD)</th>
<th>FSD (24 df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3.52 (.35)</td>
<td>3.69 (.29)</td>
<td>.17</td>
</tr>
<tr>
<td>L</td>
<td>3.69 (.49)</td>
<td>3.72 (.51)</td>
<td>.03</td>
</tr>
<tr>
<td>L-</td>
<td>3.96 (.60)</td>
<td>4.05 (.46)</td>
<td>.09</td>
</tr>
</tbody>
</table>

C = all evidence germane to the question of whether a possible defect caused the accident. \( C = (C^- + C^+) / 2 \)

L = all evidence concerning whether the defendant has any legal responsibility to the plaintiff, given that he was not the purchaser of the aircraft. \( L = (L^- + L^+) / 2 \)

L- = all evidence against the existence of a legal obligation of the manufacturer.

* indicates significance at the .05 level
CHAPTER IV
DISCUSSION

The hypothesis that structured instructions, requiring a jury to deal with the legal issues in a case one at a time, would be an aid to deliberation in a complex case was not supported. There were no significant differences among the three experimental conditions in either the number of decisions for the plaintiff nor the confidence in the correctness of the verdict. Thus the central hypothesis of the experiment was not confirmed.

There were, likewise, no significant differences in the measure taken of the way the structured and unstructured groups evaluated the evidence. Thus it can not even be asserted that, structured instructions had some effect on the way that juror's perceived the evidence even if it was not reflected in a difference in verdicts. The comparisons failed to reach significance at the .05 and the .10 levels.

There appear to be significant differences in the way the evidence was evaluated between the structured and short groups, as well as between the unstructured and short groups. Possible interpretations of these differences will be considered first. Following that, possible explanations for the failure of the central
The first issue to be considered is the differences in the way the evidence was evaluated between the short and the structured, and the short and unstructured groups. Before evaluation of differences can be undertaken, however, it must be noted that the evidence evaluation form was originally designed to answer the question of which of the experimental conditions placed more weight on evidence that was legally relevant to the case and which conditions placed more weight on emotional, but legally irrelevant factors such as the extent of the plaintiff's injuries and financial losses. The subjects are asked to rate on a Likert scale the importance of the evidence in reaching their decision. The original idea was not to break down the responses to determine how subjects responded to each category of evidence. When the predicted results did not occur, a more detailed analysis was undertaken in order to gain some clue as to what the negative results might mean. But it must be remarked that the question asked the subjects: how "important" each piece of evidence was to the decision, is somewhat ambiguous. One does not really know what "importance" means to a particular subject. A piece of evidence may be
rated important for highly idiosyncratic reasons, reasons which are related to the subject's emotions rather than to correct legal logic. Yet it is not unreasonable to hypothesize that when a subject rates pieces of evidence or argument which are highly favorable or unfavorable to an issue in the case as "important" that they are in agreement with that evidence. The expected tendency is to claim that arguments which we do not agree with are not important. As Hermann and others have found, we tend to insulate ourselves against evidence that disagrees with our position (Herman, 1977, Aronson, 1973). Thus in examining how the evidence was rated, some suggestions might be discovered as to the pattern of deliberation in the various experimental conditions. Nevertheless, the inferences from the questionnaire are not as clear as one would like because of the ambiguous nature of an evaluation of "importance."

The short group (M= 4.67) evaluated the evidence that a defect was not the cause of the accident as more important in coming to their verdicts than either the structured (M=3.73) or the unstructured (3.88) groups. Analysis by means of Fisher's significant difference test showed that the differences were significant at the .05 level: FSD (24 df)=.94, FSD (24 df)=.79 respectively. This would tend to support the contention that in the shortened version of
the case the legal issues were somewhat clearer. The question of whether a defect was the cause of the accident is a pivotal one. To make the correct legal decision a subject must be able to make a distinction between a direct and indirect cause of an accident. A headlight failure may have indirectly caused the unfortunate landing, but the failure of the headlights was in no sense a direct cause. The fact that the short group rated evidence bearing on this point significantly higher would tend to support the contention that in the shortened version of the case the issues were clearer. The arguments that the headlights were not the direct cause were arguably the most important ones in the case. The groups that heard the longer case version did not attach as much significance to these arguments, possibly because they were distracted by the multitude of other evidence and did not fully appreciate its pivotal nature. However, it must be admitted that it is not certain that the short group rated the evidence as they did because they were following the legal instructions rather than heeding a common sense notion of what was just. The lesser amount of information may have made the lack of causal relation between the defect and the accident stand out, but it cannot be stated as fact that the reason why it stood out was because the short group was applying the case law rather
than some commonsensical notion of fairness.

Both the structured (M= 3.69) and unstructured (M= 3.72) groups rated the evidence against there being a legal obligation toward Mr. Mars more highly than the short group (M=3.04). The analysis by means of Fisher's significant difference test showed that the respective differences were significant: FSD (24df)= .99, FSD (24df)= 1.07, p<.05. What this suggests is that the groups who heard the longer case did not fully understand the concept of legal responsibility. As the case was written, legal responsibility is taken to mean that there is a responsibility to redress legitimate grievances if such exist. In this case there was no legitimate grievance, though the Airlite corporation would have been responsible for damages if such had existed.

The suggestion that the groups who heard the longer case presentation may not have understood the concept of legal responsibility is supported by the fact that in the structured condition (i.e. the only condition for which data were kept as to how each issue was decided) 39 of 51 jurors said there was no legal responsibility. This suggests that even though both the structured and unstructured groups did not fully understand the complex legal instructions in the case, they were unswayed by emotional appeals and the plight of Mr. Mars. Given that
they were not moved by his appeal, it did not matter that they did not fully understand legal responsibility. Yet the evidence remains that, in fact, they were confused. They may have been more confused than the juries that heard the shorter version of the case, who rated the evidence against legal liability less important to their decisions.

The above considerations would support the contention that the case scenario presented to the structured and unstructured groups was complex enough to be somewhat confusing. If that is the case it may have been an adequate initial test of the hypothesis that structured instructions are an aid to accurate application of the law by juries in complex cases. In that event the negative results are more damaging to the central hypothesis.

The differences found through comparisons of the structured and unstructured groups with the short group in their respective evaluations of cause and liability are rather tantalizing. There is a temptation to speculate about what differences there might be between the structured and unstructured groups by using these comparisons with the short group. However, if there are no statistically significant differences between the structured and unstructured groups, it is not possible to use data demonstrating how those groups compare with a
third entity to show that there might indeed be differences between them. The short group heard a different case. It was a shorter version of the same case, but it remains to be demonstrated what effect case length might have on deliberations.

The absence of significant differences among the groups is disappointing but it may also be promising for future study, as will be considered shortly. Given that this is a new area of research, perhaps the results are not surprising. It is to be expected that a certain number of attempts might be required to perfect a methodology. In this preliminary study, it is hard to disentangle what may be limitations in the method from the possibility that the hypothesis is flawed. Yet, in that the study has yielded a number of questions for future research, ideas for improvement of the method, and that negative results have bearing on certain important legal and social issues, the effort has some benefit.

In this section hypotheses will be considered which might explain these results and, more importantly, questions will be formed which might guide future research. There are three questions that need to be considered. First, what are possible explanations for these results? The second question is related to and may often best be considered simultaneously with the first:
how might future research be conducted in order to provide a better test of the hypothesis? Thirdly, what bearing do the results have on the questions about the jury system raised by the recent insurance crisis? Negative results from this experiment have some bearing on current questions about the ability of the jury system to justly handle the increasing number of liability suits.

First, let us consider possible explanations for the results. There appear to be six obvious hypotheses. Not all are mutually exclusive; some combination of them may have been acting together. A) It may be that structured instructions are of no help whatsoever. It is possible that in complex cases, breaking the issues down and deciding them one at a time does not aid deliberation. B) In the present study it appears that the central hypothesis described above in A was not tested. The case scenario used may have been flawed. C) Perhaps it is more difficult to construct a civil trial scenario which adequately engages a jury than it is to construct a similar one for a criminal case. Almost all of the work on the validity of jury simulations has involved criminal cases. It is not obvious, on reflection, that the evidence for the adequacy of criminal trial simulations can be applied to civil cases. D) It is possible that the results of this study may be due to what social psychologists call
high objective self awareness (Wicklund, 1975). Use of a tape recorder to record the deliberations may have produced a condition where the jurors responded in ways that conform to their ideals, but not to what can be expected of actual jury behavior. E) The measures used may not have been sensitive enough to record subtle differences between groups. F) There is an implicit premise behind the present study that juries often fail to decide complex civil cases in conformance with the law. This has become a national issue, brought about by complaints from several sectors, including insurance companies, manufacturers and physicians about irrational and excessive jury awards. Perhaps juries are more sensible than we have been led to believe. If that is the case, then the results are what one would expect and the concern expressed about the jury system may be an instance of illusory correlation.

Let us consider these explanations one by one, at the same time considering ways in which research could be designed to address the unanswered questions produced by this study.

Firstly, it appears that the hypothesis advanced in the experiment has not been tested. Pilot work led the researcher to believe that there would be far more variation in verdict response from jury members. Pilot
group members maintained a lively debate, great uncertainty and variation in response to the scenario. The overwhelming rejection of the complaints of the plaintiff by the research subjects in the full scale study leaves the central question unanswered. There are two possibilities suggested by this result. First, structured instructions may be of no help to jurors in deciding complex cases. Or, secondly, this particular trial simulation, which reached a fair level of complexity, and demanded that subjects process and retain information after only aural presentation, was not complex enough for structured instructions to be needed and to prove their value. The present results might be taken to imply that a still more complex and lengthy case must be constructed to test the hypothesis.

In many ways the central issue of the case is whether a defect in the plane was the cause of the accident. All of the irrelevancies of the case were essentially means to obscure the realization that that is the central point. As stated above, an argument can be made from the results that the groups that received the shortened transcript found it easiest to get to the heart of the case. That is, short groups (M=4.67), rated the evidence that a defect was not the cause of the accident significantly higher than the structured groups (M=3.73) or the unstructured
groups (M=3.88). The Fisher's significant difference test indicated a significant difference in both comparisons: FSD (24df)=.94 and FSD (24df)=.79 respectively. The differences were significant at the .05 level. This would lead to a conclusion that the full trial presentation used was indeed more difficult than the shortened transcript. But it may not have been complex enough for differences in verdicts to begin to appear between the structured and unstructured groups. Admittedly this is speculation, but without further study the question can not be answered.

Complexity in a simulated case can be increased in two directions. The case might be made more intellectually complex, more emotionally complex, or both. To increase intellectual complexity one might add more information, more intricate legal distinctions, more arguments, more issues to be decided. Emotional complexity would be increased by making the plight of the plaintiff more severe and his character more empathetically engaging.

The results also imply a need for very extensive pilot work to refine a case to a crucial intermediate point where it is complex enough to produce variation in response, while not becoming so complex that in effect juror responses are random because no one understands the issues at all. The problem in producing such a scenario is made more difficult by the fact that the scenario created
must be capable of being run in a very limited period of time. It must clearly have a correct result, but the result can not be obvious. This is not a simple task and clearly calls for more extensive preexperimental testing than was originally envisioned.

There is a related question to the above consideration of the difficulties of creating a proper test case scenario. Almost all of the work done in the area of jury research has dealt with the problems of criminal law. There is evidence, cited above, that simulations with undergraduates are adequate approximations of courtroom behavior with such cases. But it might be questioned whether the same holds for civil trials. There is little experimental evidence to support that belief; it is not at all obvious a priori that the two types of simulation are equivalent. It may well be that there is a different empathetic situation in a civil case. In a criminal case, in essence, the problem is whether something should be taken away from defendant: liberty or life. In a civil case the question is whether something should be given to the plaintiff: damages or compensation. It is possible that many of us empathize with the prisoner on the dock and weigh his fate with some care. On the other hand the civil complainant may be seen as having created his own troubles, or at least empathy may be divided between two contestants. In the criminal case the issues are of direct
emotional immediacy: murder, rape, theft, etc. In the
civil case the actual events and responsibilities may be
more convoluted and less immediate. In the criminal case a
Type I error (to convict falsely) is abhorrent. In a civil
case a Type I error (to deny compensation where deserved)
may not seem as serious. It may well be that in the
simulation of civil cases it is more difficult to attain
the desired degree of experimental realism, in order to
engage the mock jury with the problem in the same way that
they are engaged in an actual civil trial. In reference
to the present experiment, it is possible that the
expected results were not obtained because the subjects
were not involved with the situation as a real jury would
be. In a civil case, if one can not fully feel with and
for the plaintiff then the verdict will probably go to the
defendant.

In future studies it may be necessary to use videotape
as a more immediate and emotionally forceful medium in
order to bring jury involvement to a level analogous to
that found in the courtroom. One might take a hint from
lawyers such as Peter Perlman who wrote in an article on
how to sway the civil jury that "it is incumbent upon the
trail attorney to turn the dead, inanimate stack of paper
known as the "hospital record" into a living demonstration
of pain, anxiety and depression "(Perlman, 1984).
It may be desirable in future studies to include a measure of the degree of empathy and identification with plaintiff and defendant in order to gain some better idea of how subjects are responding to the information they are presented. Information about the empathy felt for the plaintiff serves as a check on the effectiveness of the trial scenario in involving the experimental subject in a meaningful way. A measure of empathy also gives some insight into whether the jury voted on the basis of logic or emotion. If there is no empathy for the plaintiff, a vote against him may tell little about whether the jury fully analyzed the issues in the case or simply felt little concern for his plight. If, however, a simulated jury has considerable empathy for the plaintiff and still votes against him, then one can assume that they have made a fully rational decision. The same might be said, though less compellingly, if subjects have little empathy for the plaintiff and vote in his favor.

Social psychology produces another possible explanation for the experimental results. The experimental procedure required that someone be in the room to operate a tape recorder to record the jury deliberations in order to capture them for possible future research. The social-psychological research on objective self awareness leads to an interesting question in regard to the presence of
the observer and the recorder. Inadvertently, a situation was set up in which the subjects may have been made objectively self aware. The research would indicate that under those circumstances, in order to reduce dissonance between perceived self image and behavior, they would be more likely to behave according to their principles and ideals (Wicklund, 1975, Gibbons & Wicklund 1976). Or, in the more recent reformulation, even if the process of looking inward is not aversive, it leads to a continual adjustment process by which we compare our behavior to internal standards (Carver & Scheier, 1981, Scheier & Carver 1982). In any case, high self awareness tends to lead to improved performance on a variety of tasks.

In cases where juries incorrectly apply the law it might be thought of as stemming from one of two sources. First, they may well fail to comprehend the facts and the law and thus come to improper conclusions. Second, they may be tempted to apply their own version of what they think the law should be in defiance of what the law prescribes. As mentioned above, jurors are more likely to do this when confused by a complex case (Kaplan & Miller, 1978, Sue, Smith & Caldwell, 1973). A state of high self awareness might help minimize both sources of error. Objective self awareness might have led subjects to listen and deliberate more carefully, thus working harder than a jury ordinarily
might. It has been repeatedly established that high self awareness increases the care with which a task is accomplished. Second, knowing that they were being taped, a temptation on the part of subjects to side with the defendant out of pity might well be minimized. There has been a good deal of attention in the media about the crisis in the insurance industry and problems of large jury awards. It is reasonable to believe that a good number of the subjects were well aware that they "should" not let their emotions get in the way of an objective verdict. A state of high self awareness may well have made it less likely that they would do so.

In the future it may be best not to tape the mock jury deliberations and to have no outside observer in the room in order to avoid this possible interference with the experimental manipulation.

There is another possible explanation for the results. Structured instructions would doubtlessly make it necessary to debate a case more systematically and thoroughly. Examining the case in a more systematic fashion may well lead to a more fully considered approach to both sides, in itself causing some confusion.

There may be two kinds of jury confusion. One type may be evoked by extreme complexity of evidence and strength of emotional appeal. The juror may be left with a global
impression of the events he has witnessed which may not be at all in conformance with the law or the evidence. And worse, this painful state of confusion might leave him unwilling or unable to consider the evidence in a multifaceted way. Having achieved some tenuous command of what he had just witnessed, the juror may be quite unwilling to consider other possibilities to his own construal of the evidence. This might be thought of, to borrow a term from Shapiro's study of neurotic style, as the "hysterical fallacy" in jury deliberation (1965). Shapiro said of hysterical cognition that it is "global, relatively diffuse, and lacking in sharpness, particularly in sharp detail." He adds that for the hysterical person "the hunch or the impression is the final, conscious cognitive product (1965)."

In some cases a jury might enter deliberation seeking only to confirm their global prejudices, determined to gloss over differences and ignore distinctions that get in the way of consensus. This kind of confusion might be decreased by forcing the jury to engage in a systematic deliberation procedure, such as the one in this study.

There may be a second type of confusion, however. This, to again borrow a term from Shapiro, might be thought of as leading to the "obsessive fallacy" in deliberation. Shapiro describes the obsessive's process of attention as
being "markedly limited in both mobility and range...And some aspects of the world are simply not to be apprehended by sharply focused and concentrated attention (1965)."

Structured deliberation might encourage getting bogged down in details of argument to the point where the central focus of the case is lost sight of in the welter of detail. A structured instruction set forces consideration of all arguments systematically. If the obsessive model is apt, this can possibly lead to an impulsive decision after lengthy rumination. While these thoughts are highly speculative, it seems worth considering that perhaps systematic deliberation is a two edged sword. The one jury that voted solidly for the plaintiff was in the structured condition. While the verdict patterns between jury groups in the structured and unstructured conditions do not reveal a significant difference, the fact that the only jury to vote for the plaintiff was in the structured condition is worth brief consideration. It is possible that structured deliberation can sometimes hinder deliberation until a case reaches a certain level of complexity. When structured instructions are used on a case that is too simple for them to be appropriate, they may encourage preoccupation with details of the evidence and argument that may obscure attaining a sense of the case as a whole. If this is the case, the issue of whether
structured instructions are helpful in jury deliberations would need to take into account two factors. First, structured instructions might aid deliberation by making sure that all of the evidence and all of the law are systematically considered. But, secondly, it is also possible that under certain circumstances they may hinder deliberation by discouraging the attainment of a global view of the case. If that is true, the problem of whether structured instructions should be used in the legal system might be very complex. In support of the notion that structured instructions may in some cases not help, when comparing jury verdicts, using the individual subject as the unit of measurement, there was a significant difference at the .05 level between the structured and short groups but none between the unstructured and short groups. While it is problematic, to attempt to draw conclusions about differences between groups by comparing them with a third entity, it would appear that structured instructions made the case somewhat more difficult for the structured group. The unstructured group, in rendering verdicts, behaved like the short group, which had a simpler task. The structured group behaved in a way significantly different from the short group.

There is another possible flaw in the experimental design that needs to be considered. In this experiment
subjects were required to decide to either give the plaintiff the money he asked for or give nothing. There was no possibility to temporize and give a small award. The forced choice may have removed a great deal of sensitivity from the measure. Because the jurors were not allowed to strike a compromise, they may have been stricter than they would ordinarily be inclined to be. It is possible that if jurors had been allowed to give various sizes of awards, differences between groups might have appeared.

A number of scientific studies, such as those cited in the introduction have questioned the accuracy and fairness of jury deliberations. Similar doubts are increasingly voiced in the legal community and the popular press about the ability of juries to reach conclusions in correct conformance with the law and the evidence. The year of 1986 saw a kind of high water mark of concern over the tort system marked by a perceived insurance crisis where horror stories circulated of "price hikes of up to 1000% and abrupt, unprovoked cancellation of policies" (Farrell, 1987). Some claimed the crisis was manufactured by the insurance industry to raise rates. Yet as US Assistant Attorney General Richard K. Willard stated: "We don't think there is credible evidence that the liability crisis is a hoax. The insurance industry is the middleman" (Gatty
Men like Blair Childs, director of the American Tort Reform Association said, "We're in trouble because we have moved so far away from a court system based on fault. We've moved to a compensation system, with the concept that somebody must pay regardless." (Gatty, 1987) In response to the problem three dozen states made minor changes in their liability laws (Farrell, 1987). The efforts for reform of the civil justice system have not yet abated. An initiative to limit jury awards was on the ballot last election in Montana and was approved.

At first glance the results from this experiment would tend to question whether the current cynicism about jury awards is warranted. The results are somewhat reassuring. In this study, using a convoluted case with complex instructions, involving direct emotional appeal to the sympathies of the jury, a total of 25 of the 27 groups who heard the case voted solidly for the defendant. In 3 juries there was a minority of 1 who voted for the plaintiff, in a single jury there was a minority of 2. Given the evidence that the majority will usually sway the dissenters, one can probably assume that given further deliberation the verdict would have been unanimous for the defendant. The results raise the question of whether a great deal of presently expressed concern about the jury system is justified. Christopher Farrell, a correspondent
for *Business Week* wrote in May of 1987 that "Cool analysis is discrediting last year's horror stories about an epidemic of multimillion-dollar jury awards for relatively little cause." (Farrell, 1987) He cites an American Enterprise Institute study of 359 cases from 1982-85 mostly involving product liability showing that in that sampling punitive damages represented "insignificant" amounts. Perhaps some of the current concern about the legal system represents an instance of illusory correlation as described by Chapman (Chapman, 1967). The presence of a few confirming instances may be leading the public to conclude that there is a grave structural problem where none exists.

One jury was evenly split, one voted solidly for the plaintiff. It is impossible to say in what direction the split jury would have voted. Thus it is possible that 2 of the 27 juries would have voted for the plaintiff. That would mean that 7.4% would have made a decision not in conformance with the law and the evidence. On the one hand it is reassuring, given the concerns that have been expressed about the jury system that there weren't more errors. Yet on further consideration justice that fails 7.4% of the time or even 3.7% is worrisome. The concept of an error factor in human behavior seems benign until one begins to envision it depriving citizens of their
property. The experiment obviously does not begin to give an estimate of what the error factor in actual jury deliberations might be. But the point, rather, seems to be that here one has what appears to most juries to be an open and shut case. And yet one jury was evenly split and another voted solidly opposite to majority opinion.

One of the conclusions the present study might suggest is that instances in which a jury is swayed by sympathy for the plaintiff to give inappropriate awards may be rather infrequent. And this would not be a trivial result.

If it is the case that jury mistakes in this area are infrequent, then to study the phenomenon it would be necessary to do a number of things. Much larger sample sizes would be needed. Given a very low base line phenomena a large sample is needed to detect differences between groups. Given that the sample sizes might need to be so large as to be impractical, if groups are the unit of observation, the inclination might be to declare the unit of observation to be the individual in the early stages of experimentation. While it does not necessarily follow that differences in individuals will be transferred into group differences once individuals are subjected to group process, individual perceptions are the raw materials with which the group works. If there are no differences in individual perceptions, there is no need to
seek them through group process.

Further, it might be suggested that a more homogeneous group should be sought in order to decrease the error variance. Thus external validity would be sacrificed at an early stage of research in order to first try to establish a model to be carried into later research with greater real world application. Given that the subject pool in this experiment was fairly homogeneous already, this may not be a very practical suggestion.

As a final thought, there is the consideration of vulnerabilities in the legal system suggested by these results. On the one hand it is suggested that there may be greater abilities on the part of jurors to sort through diversions, digressions and irrelevancies than many have recently been led to believe. In this study, a case designed to confuse, did not do so for the vast majority of subjects. That is rather good news. Yet one jury voted solidly for the plaintiff. And one jury was split. This would seem to suggest that it is worthwhile to study ways to reduce jury confusion. The purpose of this study and others like it is to attempt to find ways to decrease vulnerability to juror error. But in thinking about the results there is a radical question that suggests itself. Doesn't this kind of effort on the part of psychology ignore what may be the greatest source of juror error? Our
tradition of law makes the lawyer an advocate, responsible to little other than his advocacy. He must present the best case possible. He may not lie, but he does not need to ask himself what the truth may be. His sole responsibility is to represent his client as well as he can. As Sir Edward Marshall Hall expressed it: "A lawyer and an actor are akin. It is true that I have no mask, I have no black cloth and I have no floodlights to help bring illusion; but out of the miseries and the joys and experience of men, I must create an atmosphere of living reality so that it may be felt and understood by others, for this is advocacy." (Perlman, 1981) Given what we know about the difficulties subjects have in processing complex, emotionally charged information, one is not at all confident that justice is assured by having two great actors engage in a duel of thespian skill. Perhaps lawyers might be taught that they have an obligation to truth as well as to clients. It would be a different way of looking at the law and their work. It is not a likely suggestion or seriously advanced as a possibility. But in considering the problem of jury perception of complex materials, at times one wonders how far justice has advanced from trial by combat. Yet given the fact that the legal tradition is not likely to change in the near future, studies such as the present one seem vital to determine how best to aid
the juror in untangling the spells that lawyers attempt to weave.

**SUMMARY**

In summary, the hypothesis of the experiment that structured instructions can be of help to jurors as they attempt to apply the law to a complex case was not supported. There were no significant differences between the structured and unstructured groups. The results may be due to the fact that the hypothesis is not true or that the scenario was not complex enough to warrant the use of structured instructions. It is even possible that structured instructions may have some hindering effects on deliberation which might cancel their positive benefits. It is possible that, inadvertently, a condition of high self awareness was created in the jurors, thus making them more conscientious than they would ordinarily be. It is also possible that the jurors were not sufficiently engaged by the scenario. The measures may have been lacking in sensitivity or perhaps contrary to popular belief jurors are better able to apply the law than we have been led to believe. All of these questions can only be resolved by further research.
REFERENCES


Bower, G. H., Black, J.B., & Turner, T. J. (1979), Scripts in memory for text. *Cognitive Psychology*, 11, 177-220


Farrell, Christopher. (1987). The crisis is over- But insurance will never be the same, *Business Week*, #3000, 122-23


Appendix A

Airlite, a manufacturer of small airplanes, guaranteed that it would pay for any injury to the buyer of one of its planes which resulted from the operation of an Airlite aircraft, if the reason for the damage was the negligence of Airlite. Roger Hanson lent his friend Frank Mars his newly purchased Airlite plane for the weekend. Frank was a careful pilot and had never been in an accident. Just after takeoff, he noticed that the plane's headlights did not appear to be working. He decided to turn back to the airport and have them repaired before dark. However, as Frank was approaching the runway, the plane was caught by a sudden gust of wind that sent it crashing into the ground.

Mr. Mars broke his arm and fractured his skull. He was in the hospital for six weeks and when he got out he found that he did not have full use of his left hand because of the injury and he suffered occasional dizzy spells on account of the blow to his head. Frank Mars was a heavy equipment operator and thus found himself unable to work. His hospital bills were far beyond his ability to pay. His car was repossessed and he fell behind in payments on his house. Working part time as a box boy, the only work he could get, and with his wife working as a secretary, they could barely support their five children.
Thus Frank Mars sued Airlite for the cost of his injuries and the loss of his livelihood, asking for money for retraining in another occupation.

Frank contended that if it were not for the failed headlight on the airplane he would not have tried to make a landing when he did and would not have crashed. Thus, according to his attorney, the failed headlight was the cause of the accident.

Airlite contended that the cause of the accident was the gust of wind. No structural defect in the aircraft had caused the crash. The Airlite attorney asked, "If poor Mr. Mars has turned back because he forgot his lunch, would we then be responsible? If you answer no, then you can not hold us responsible in this accident. There is nothing that we did or did not do that caused this unfortunate mishap."

Airlite further claimed that their guarantee was to the purchaser of the aircraft, Roger Hansen, and that under state law they had little obligation to Mr. Mars who did not actually buy the plane from them. The lawyer for Airlite argued that since Mr. Hanson lent the plane, it was his responsibility if the plane was defective.

The lawyer for Mr. Mars argued that although a manufacturer's legal obligation was primarily to the original buyer, the law was never intended to mean that if
you weren't the original buyer, the manufacturer had no responsibility at all. In this case, Mr. Hanson had lent Mr. Mars a new plane and had no reason to believe that there was a defect. Therefore he had no responsibility.

The lawyer for Airlite argued that Frank Mars should have inspected the headlights before he took off.

Frank's attorney countered that headlight inspection was not part of the standard pre-flight check and that it was only by accident that Frank had happened to pull the light switch and notice the light problem once he was in the air.

Frank Mars' attorney also argued that the type of headlight installed on the model 1200 Airlite, which Frank was flying, was a new experimental type of aviation light which had a reputation for being unreliable. He also tried to prove that after the accident, faulty soldering was found in one of the light connections. A former employee of Airlite was testified that in his experience the electrical work on Airlite planes was often faulty.

The lawyer for the Airlite company, however, brought out the fact that the employee who claimed that Airlite work was often faulty had been fired from the company for drug use and had left making threats to get even.

In defense of Airlite, their lawyer produced witnesses who testified that the headlights on the model 1200 were
the same as those used on 5 other models of small plane and that there had been few problems reported. Expert witnesses also stated that the claim that the soldering connections on the headlights had been faulty could not be proven, because the plane had been so badly damaged in the accident that there was no way anyone could tell if the connection had been badly wired by Airlite.

The attorney for Mr. Mars argued that although there had not been a large number of failures of the new type headlights on the Airlite plane, a sophisticated statistical analysis of the failure rate revealed that there was a high probability that the headlights had a design defect.

The Airlite attorneys ridiculed the statistical analysis of Mr. Mars' attorney, saying you could make any point you wanted with statistics. He quoted the saying that "there were three kinds of lies: lies, damn lies and statistics."

The Airlite attorney further contended that Mr. Hanson, the original owner of the plane, was an inexperienced pilot known for his rough landings. He very well could have jolted the plane so badly in the week he owned it, that he damaged the headlights before he had lent it to his unfortunate friend.

The lawyer for Mr. Mars argued that the plane was barely a week old. Even if there had been some rough landings, it
surely should have been built to withstand a bump or two without wiring coming loose.

LEGAL PRINCIPLES:

1) If the nature of a product is such that it is reasonably certain to place life and limb in danger, the manufacturer is legally responsible for all damages to the purchaser due to the manufacturer's negligence.

2) The manufacturer's primary responsibility is to the purchaser of his product. He has a lesser legal responsibility to those who come to use his products without having bought them from him or his agents. If the product is acquired by theft, there is no responsibility. If it is borrowed, the lender may have equal responsibility for the condition of the product.

3) A person or company shall be considered negligent if they are directly responsible for an injury to another person who had reason to expect that they would take precautions to prevent that injury.

4) A person can be considered directly responsible for someone else's injury if their act made that injury, in the view of a reasonable and prudent man, nearly unavoidable.
Appendix B

Airlite, was a manufacturer of small airplanes. Roger Hanson lent his friend Frank Mars his newly purchased Airlite plane for the weekend. Just after takeoff, Frank Mars noticed that the plane's headlights did not appear to be working. He decided to turn back to the airport and have them repaired before dark. However, as Frank was approaching the runway, the plane was caught by a sudden freak gust of wind that sent it crashing into the ground.

Mr. Mars broke his arm and fractured his skull. He was in the hospital for six weeks and when he got out he found that he did not have full use of his left hand because of the injury and he suffered occasional dizzy spells an account of the blow to his head. Frank Mars was a heavy equipment operator and thus found himself unable to work. Thus Frank Mars sued Airlite for the cost of his injuries and the loss of his livelihood, asking for money for retraining in another occupation.

Frank contended that if it were not for the failed headlight on the airplane he would not have tried to make a landing when he did and thus would not have crashed. Thus, according to his attorney, the failed headlight was the cause of the accident.

Airlite contended that the cause of the accident was the
gust of wind. No structural defect in the aircraft was responsible for the accident. The Airlite attorney asked, "If poor Mr. Mars has turned back because he forgot his lunch, would we then be responsible? If you answer no, then you can not hold us responsible in this accident. There is nothing that we did or did not do that caused this unfortunate mishap."

The lawyer for Frank Mars produced an expert witness who testified that faulty soldering was found in one of the light connections of the plane after the accident. A former employee of Airlite also testified that the electrical work on Airlite planes was often faulty.

The lawyer for the Airlite company, however, brought out the fact that the employee who claimed that Airlite work was often faulty had been fired from the company for drug use and might be trying to get even. Airlite also produced expert witnesses of its own who testified that the plane was so badly damaged in the accident that it was difficult to tell if the light connection had been badly wired by Airlite.

Airlite further claimed that their guarantee was to the purchaser of the aircraft, Roger Hansen, and that under state law they had little obligation to Mr. Mars who did not actually buy the plane from them. The lawyer for Airlite argued that since Mr. Hanson lent the plane, it
was his responsibility if the plane was defective.

The lawyer for Mr. Mars argued that although a manufacturer's legal obligation was primarily to the original buyer, the law was never intended to mean that if you weren't the original buyer, the manufacturer had no responsibility at all. In this case, Mr. Hanson had lent Mr. Mars a new plane and had no reason to believe that there was a defect. Therefore he had no responsibility.

LEGAL PRINCIPLES:

1) If the nature of a product is such that it is reasonably certain to place life and limb in danger, the manufacturer is legally responsible for all damages to the purchaser due to the manufacturer's negligence.

2) The manufacturer's primary responsibility is to the purchaser of his product. He has a lesser legal responsibility to those who come to use his products without having bought them from him or his agents. If the product is acquired by theft, there is no responsibility. If it is borrowed, the lender may have equal responsibility for the condition of the product.

3) A person or company shall be considered negligent if they are directly responsible for an injury to another person who had reason to expect that they would take precautions to prevent that injury.

4) A person can be considered directly responsible for
someone else's injury if their act made that injury, in
the view of a reasonable and prudent man, nearly
unavoidable.
APPENDIX C

Structured Instructions (read to the jurors)

USE THE LEGAL PRINCIPLES STATED IN THE CASE TO DECIDE THE FOLLOWING QUESTIONS:

1) Does Airlite have a legal responsibility to Frank Mars?
   YES___ NO___

2) Did Airlite cause the defect in the headlights?
   YES___NO___

3) Was that defect responsible for the accident?
   YES___NO___

   If you answer "yes" to all of the above questions, that is, if you decide that: 1) Airlite is legally responsible to Mr. Mars, 2) the plane was defectively made by Airlite, and 3) the defect was directly responsible for the accident, then Airlite was negligent and you must decide in favor of Mr. Mars.

WHAT IS YOUR VERDICT?

(please place a check mark by your verdict)

I DECIDE IN FAVOR OF MR. MARS_____  
I DECIDE IN FAVOR OF AIRLITE CO._____
Unstructured Instructions
(read to the jurors)

USE THE LEGAL PRINCIPLES STATED IN THE CASE TO DECIDE THE FOLLOWING QUESTIONS:

Does Airlite have a legal responsibility to Frank Mars?
Did Airlite cause the defect in the headlights?
Was that defect responsible for the accident?

If you answer "yes" to all of the above questions, that is, if you decide that: 1) Airlite is legally responsible to Mr. Mars, 2) the plane was defectively made by Airlite, and 3) the defect was directly responsible for the accident, then Airlite was negligent and you must decide in favor of Mr. Mars.

WHAT IS YOUR VERDICT?

(please place a check mark by your verdict)

I DECIDE IN FAVOR OF MR. MARS____

I DECIDE IN FAVOR OF AIRLITE CO._____
APPENDIX D

Please answer the questions below as carefully as you can. Your answers will be compared to those of actual jurors. The information you give is vital to the success of the experiment. I very much appreciate your cooperation.

Below you will find the a list of statements made in the case you just heard. Below each statement you will also find a rating scale. Please rate how important each piece of evidence was in arriving at your verdict.

For instance, if think a fact was not important in deciding the case the way you did, you would circle the "1" on the scale.

If you felt the fact was extremely important in arriving at your verdict, you would circle the "5" on the scale.

For example, if you felt that it was a very important piece of information in helping you reach your verdict that "Frank Mars was a careful pilot," you would rate that statement "4." And you would circle the "4" on the scale below the statement, just like on the scale below.

1  2  3  4  5
not a little somewhat very extremely important important

*** Thank you for your help. ***

1) Airlite, guaranteed that it would pay for any injury to the buyer of one of its planes which resulted from the operation of an Airlite aircraft, if the reason for the damage was the negligence of Airlite.

1  2  3  4  5
not a little somewhat very extremely important important

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
2) Roger Hanson lent his friend Frank Mars his newly purchased Airlite plane for the weekend.

3) Frank was a careful pilot and had never been in an accident.

4) Just after takeoff, he noticed that the plane's headlights did not appear to be working.

5) He decided to turn back to the airport and have them repaired before dark.

6) However, as Frank was approaching the runway, the plane was caught by a sudden freak gust of wind that sent it crashing into the ground.
7) Mr. Mars broke his arm and fractured his skull.

8) He was in the hospital for six weeks and when he got out he found that he did not have full use of his left hand because of the injury and he suffered occasional dizzy spells an account of the blow to his head.

9) Frank Mars was a heavy equipment operator and thus found himself unable to work.

10) His hospital bills were far beyond his ability to pay.
11) His car was repossessed and he fell behind in payments to his house.

1 2 3 4 5
not a little somewhat very extremely important

12) Working part time as a box boy, the only work he could get, and with his wife working as a secretary, they could barely support his five children.

1 2 3 4 5
not a little somewhat very extremely important

13) Frank contended that if it were not for the failed headlight on the airplane he would not have tried to make a landing when he did and thus would not have crashed.

1 2 3 4 5
not a little somewhat very extremely important

14) Airlite contended that the cause of the accident was the gust of wind.

1 2 3 4 5
not a little somewhat very extremely important

15) No structural defect in the aircraft was responsible for the accident.

1 2 3 4 5
not a little somewhat very extremely important
16) Airlite further claimed that their guarantee was to the purchaser of the aircraft, Roger Hansen, and that under state law they had a minimal obligation to Mr. Mars who did not actually buy the plane from them. The lawyer for Airlite argued that since Mr. Hanson lent the plane, it was his responsibility if the plane was defective.

17) The lawyer for Mr. Mars argued that although a manufacturer's legal obligation was primarily to the original buyer, the law was never intended to mean that if you weren't the original buyer, the manufacturer had no responsibility at all. In this case, Mr. Hanson had lent Mr. Mars a new plane and had no reason to believe that there was a defect. Therefore he had no responsibility.

18) Airlite claimed that Frank Mars should have inspected the headlights before he took off.

19) Frank's attorney countered that headlight inspection was not part of the standard pre-flight check.
20) Frank Mars' attorney argued that the type of headlight installed on the model 1200 Airlite, which Frank was flying, was a new experimental type of aviation light which had a reputation for being unreliable.

21) Mr. Mars' attorney tried to prove that after the accident, faulty soldering was found in one of the light connections.

22) Expert witnesses also stated that the claim that the soldering connections on the headlights had been faulty could not be proven, because the plane had been so badly damaged in the accident that there was no way anyone could tell if the connection had been badly wired by Airlite.

23) A former employee of Airlite was produced as a witness who testified that electrical work on Airlite planes was often faulty.
24) The lawyer for the Airlite company, however, brought out the fact that the employee who claimed that Airlite work was often faulty had been fired from the company for drug use and had left making threats to get even.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>not</td>
<td>a little</td>
<td>somewhat</td>
<td>very</td>
<td>extremely important</td>
</tr>
</tbody>
</table>

25) In defense of Airlite, their lawyer produced witnesses who testified that the headlights on the model 1200 were the same as those used on 5 other models of small plane and that there had been few problems reported.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>not</td>
<td>a little</td>
<td>somewhat</td>
<td>very</td>
<td>extremely important</td>
</tr>
</tbody>
</table>

26) The attorney for Mr. Mars argued that although there had not been a large number of failures of the new type headlights on the Airlite plane, a sophisticated statistical analysis of the failure rate revealed that there was a high probability that the headlights had a design defect.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>not</td>
<td>a little</td>
<td>somewhat</td>
<td>very</td>
<td>extremely important</td>
</tr>
</tbody>
</table>

27) The Airlite attorneys ridiculed the statistical analysis of Mr. Mars attorney, saying you could make any point you wanted with statistics. He quoted the saying that "there were three kinds of lies: lies, damn lies and statistics."

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>not</td>
<td>a little</td>
<td>somewhat</td>
<td>very</td>
<td>extremely important</td>
</tr>
</tbody>
</table>
28) The Airlite attorney further contended that Mr. Hanson, the original owner of the plane, was an inexperienced pilot known for his rough landings. He very well could have jolted the plane so badly in the week he owned it, that he damaged the headlights himself before he had lent it to his unfortunate friend.

1 2 3 4 5

not a little somewhat very extremely important important

29) The lawyer for Mr. Mars argued that the plane was barely a week old. Even if there had been some rough landings, it surely should have been built to withstand a bump or two without wiring coming loose.

1 2 3 4 5

not a little somewhat very extremely important important

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
APPENDIX E

Please answer the questions below as carefully as you can. Your answers will be compared to those of actual jurors. The information you give is vital to the success of the experiment. I very much appreciate your cooperation.

Below you will find the a list of statements made in the case you just heard. Below each statement you will also find a rating scale. Please rate how important each piece of evidence was in arriving at your verdict.

For instance, if think a fact was not important in deciding the case the way you did, you would circle the "1" on the scale.

If you felt the fact was extremely important in arriving at your verdict, you would circle the "5" on the scale.

For example, if you felt that it was a very important piece of information in helping you reach your verdict that "Frank Mars was a careful pilot," you would rate that statement "4." And you would circle the "4" on the scale below the statement, just like on the scale below.

1 2 3 4 5
not a little somewhat very extremely important

-----------------------------------------------

*** Thank you for your help. ***

1) Airlite, guaranteed that it would pay for any injury to the buyer of one of its planes which resulted from the operation of an Airlite aircraft, if the reason for the damage was the negligence of Airlite.

1 2 3 4 5
not a little somewhat very extremely important

-----------------------------------------------

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
2) Roger Hanson lent his friend Frank Mars his newly purchased Airlite plane for the weekend.

3) Just after takeoff, he noticed that the plane's headlights did not appear to be working.

4) He decided to turn back to the airport and have them repaired.

5) However, as Frank was approaching the runway, the plane was caught by a sudden gust of wind that sent it crashing into the ground.

6) Mr. Mars broke his arm and fractured his skull.
7) He was in the hospital for six weeks and when he got out he found that he did not have full use of his left hand because of the injury and he suffered occasional dizzy spells an account of the blow to his head.

not a little somewhat very extremely important

8) Frank Mars was a heavy equipment operator and thus found himself unable to work.

not a little somewhat very extremely important

9) Frank contended that if it were not for the failed headlight on the airplane he would not have tried to make a landing when he did and thus would not have crashed.

not a little somewhat very extremely important

10) Airlite contended that the cause of the accident was the gust of wind.

not a little somewhat very extremely important
11) No structural defect in the aircraft was responsible for the accident.

1  2  3  4  5
not a little somewhat very extremely important

12) Airlite further claimed that their guarantee was to the purchaser of the aircraft, Roger Hansen, and that under state law they had little obligation to Mr. Mars who did not actually buy the plane from them. The lawyer for Airlite argued that since Mr. Hanson lent the plane, it was his responsibility if the plane was defective.

1  2  3  4  5
not a little somewhat very extremely important

13) The lawyer for Mr. Mars argued that although a manufacturer's legal obligation was primarily to the original buyer, the law was never intended to mean that if you weren't the original buyer, the manufacturer had no responsibility at all. In this case, Mr. Hanson had lent Mr. Mars a new plane and had no reason to believe that there was a defect. Therefore he had no responsibility.

1  2  3  4  5
not a little somewhat very extremely important

14) The lawyer for Frank Mars produced an expert witness who testified that faulty soldering was found in one of the light connections of the plane after the accident.

1  2  3  4  5
not a little somewhat very extremely important
15) A former employee of Airlite also testified that the electrical work on Airlite planes was often faulty.

not a little somewhat very extremely important

16) The lawyer for the Airlite company, however, brought out the fact that the employee who claimed that Airlite work was often faulty had been fired from the company for drug use and might be trying to get even.

not a little somewhat very extremely important

17) Airlite also produced expert witnesses of its own who testified that the plane was so badly damaged in the accident that it was difficult to tell if the light connection had been badly wired by Airlite.

not a little somewhat very extremely important

*** Thank you for your participation. ***
Please indicate on a scale from 1 to 5, how certain you are that the verdict you reached is the correct one. Circle the response that seems the most true of how confident you are about your decision.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not very</td>
<td>Somewhat</td>
<td>Extremely</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Confident</td>
<td>Confident</td>
<td>Confident</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Structured Group - Unstructured Group

<table>
<thead>
<tr>
<th></th>
<th>Structured Mean (SD)</th>
<th>Unstructured (Pooled SD) Mean (SD)</th>
<th>F (1,16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ir</td>
<td>2.58 (.40)</td>
<td>2.64 (.59)</td>
<td>.05</td>
</tr>
<tr>
<td>R</td>
<td>3.46 (.26)</td>
<td>3.47 (.44)</td>
<td>.01</td>
</tr>
<tr>
<td>Ir/R</td>
<td>.76 (.10)</td>
<td>.78 (.14)</td>
<td>.05</td>
</tr>
<tr>
<td>D</td>
<td>3.11 (.30)</td>
<td>2.85 (.47)</td>
<td>1.92</td>
</tr>
<tr>
<td>D-</td>
<td>3.03 (.35)</td>
<td>2.83 (.59)</td>
<td>.74</td>
</tr>
<tr>
<td>D+</td>
<td>3.20 (.58)</td>
<td>2.88 (.48)</td>
<td>.53</td>
</tr>
<tr>
<td>C</td>
<td>3.52 (.35)</td>
<td>3.69 (.29)</td>
<td>1.28</td>
</tr>
<tr>
<td>C-</td>
<td>3.73 (.18)</td>
<td>3.88 (.24)</td>
<td>2.29</td>
</tr>
<tr>
<td>C+</td>
<td>3.31 (.64)</td>
<td>3.50 (.39)</td>
<td>.60</td>
</tr>
<tr>
<td>L</td>
<td>3.69 (.49)</td>
<td>3.72 (.51)</td>
<td>.02</td>
</tr>
<tr>
<td>L-</td>
<td>3.96 (.60)</td>
<td>4.05 (.46)</td>
<td>.13</td>
</tr>
<tr>
<td>L+</td>
<td>3.41 (.62)</td>
<td>3.38 (.70)</td>
<td>.01</td>
</tr>
</tbody>
</table>

Ir = average of the scores measuring the importance of irrelevant evidence (evidence concerning the seriousness of Mr. Mars' injuries and the extent of his financial losses).

R = average of the evaluations of the importance of all other case evidence.

Ir/R = the ratio between the two above measures.

D = all evidence regarding whether there was a defect in the plane. D = (D- + D+)/2.

D- = all evidence against the existence of a defect.

D+ = all evidence supporting the existence of a defect.

C = all evidence germane to the question of whether a possible defect caused the accident. C = (C- + C+)/2.

C- = all evidence against the proposition that a defect in the plane caused the accident.

C+ = all evidence supporting the claim that a defect in the plane caused the accident.

L = all evidence concerning whether the defendant has any legal responsibility to the plaintiff, given that he was not the purchaser of the aircraft. L = (L- + L+)/2.

L- = all evidence against the existence of a legal obligation of the manufacturer.

L+ = all evidence supporting the existence of a legal liability.
APPENDIX H

Unstructured Group - Short Group

<table>
<thead>
<tr>
<th>Unstructured</th>
<th>Short (Pooled SD)</th>
<th>F (1,16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Ir</td>
<td>2.64 (.59)</td>
<td>2.49 (.59)</td>
</tr>
<tr>
<td>R</td>
<td>3.47 (.44)</td>
<td>3.41 (.30)</td>
</tr>
<tr>
<td>Ir/R</td>
<td>.78 (.14)</td>
<td>.73 (.15)</td>
</tr>
<tr>
<td>D</td>
<td>2.85 (.47)</td>
<td>2.80 (.35)</td>
</tr>
<tr>
<td>D-</td>
<td>2.83 (.59)</td>
<td>3.05 (.20)</td>
</tr>
<tr>
<td>D+</td>
<td>2.88 (.48)</td>
<td>2.55 (.61)</td>
</tr>
<tr>
<td>C</td>
<td>3.69 (.29)</td>
<td>3.86 (.20)</td>
</tr>
<tr>
<td>C-</td>
<td>3.88 (.24)</td>
<td>4.67 (.21)</td>
</tr>
<tr>
<td>C+</td>
<td>3.50 (.39)</td>
<td>3.06 (.57)</td>
</tr>
<tr>
<td>L</td>
<td>3.72 (.51)</td>
<td>3.04 (.46)</td>
</tr>
<tr>
<td>L-</td>
<td>4.05 (.46)</td>
<td>2.98 (.60)</td>
</tr>
<tr>
<td>L+</td>
<td>3.38 (.70)</td>
<td>3.09 (.90)</td>
</tr>
</tbody>
</table>

Ir = average of the scores measuring the importance of irrelevant evidence (evidence concerning the seriousness of Mr. Mars' injuries and the extent of his financial losses).

R = average of the evaluations of the importance of all other case evidence.

Ir/R = the ratio between the two above measures.

D = all evidence regarding whether there was a defect in the plane. D = (D- + D+) / 2.
D- = all evidence against the existence of a defect.
D+ = all evidence supporting the existence of a defect.

C = all evidence germane to the question of whether a possible defect caused the accident. C = (C- + C+)/2
C- = all evidence against the proposition that a defect in the plane caused the accident.
C+ = all evidence supporting the claim that a defect in the plane caused the accident.

L = all evidence concerning whether the defendant has any legal responsibility to the plaintiff, given that he was not the purchaser of the aircraft. L = (L- + L+)/2
L- = all evidence against the existence of a legal obligation of the manufacturer.
L+ = all evidence supporting the existence of a legal liability.
## APPENDIX I
### Structured Group- Short Group

<table>
<thead>
<tr>
<th>Structured Mean (SD)</th>
<th>Short Mean (SD)</th>
<th>(Pooled SD) F (1,16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ir 2.58 (.40)</td>
<td>2.49 (.59)</td>
<td>(.50) .14</td>
</tr>
<tr>
<td>R 3.46 (.26)</td>
<td>3.41 (.30)</td>
<td>(.28) .13</td>
</tr>
<tr>
<td>Ir/R .76 (.10)</td>
<td>.73 (.15)</td>
<td>(.13) .09</td>
</tr>
<tr>
<td>D 3.11 (.30)</td>
<td>2.80 (.35)</td>
<td>(.32) 4.02</td>
</tr>
<tr>
<td>D- 3.03 (.35)</td>
<td>3.05 (.20)</td>
<td>(.29) .03</td>
</tr>
<tr>
<td>D+ 3.20 (.58)</td>
<td>2.55 (.61)</td>
<td>(.60) 5.27*</td>
</tr>
<tr>
<td>C 3.52 (.35)</td>
<td>3.86 (.20)</td>
<td>(.29) 6.62*</td>
</tr>
<tr>
<td>C- 3.73 (.18)</td>
<td>4.67 (.21)</td>
<td>(.20) 103.35*</td>
</tr>
<tr>
<td>C+ 3.31 (.64)</td>
<td>3.06 (.57)</td>
<td>(.60) .77</td>
</tr>
<tr>
<td>L 3.69 (.61)</td>
<td>3.04 (.72)</td>
<td>(.65) 5.53*</td>
</tr>
<tr>
<td>L- 3.96 (.60)</td>
<td>2.97 (.60)</td>
<td>(.60) 11.92*</td>
</tr>
<tr>
<td>L+ 3.41 (.62)</td>
<td>3.09 (.90)</td>
<td>(.77) .81</td>
</tr>
</tbody>
</table>

Ir= average of the scores measuring the importance of irrelevant evidence (evidence concerning the seriousness of Mr. Mars' injuries and the extent of his financial losses).

R= average of the evaluations of the importance of all other case evidence.

Ir/R= the ratio between the two above measures.

D= all evidence regarding whether there was a defect in the plane. D= (D- + D+)/2.

D- = all evidence against the existence of a defect.

D+ = all evidence supporting the existence of a defect.

C= all evidence germane to the question of whether a possible defect caused the accident. C= (C- + C+)/2

C- = all evidence against the proposition that a defect in the plane caused the accident.

C+ = all evidence supporting the claim that a defect in the plane caused the accident.

L= all evidence concerning whether the defendant has any legal responsibility to the plaintiff, given that he was not the purchaser of the aircraft. L= (L- + L+)/2

L- = all evidence against the existence of a legal obligation of the manufacturer.

L+ = all evidence supporting the existence of a legal liability.