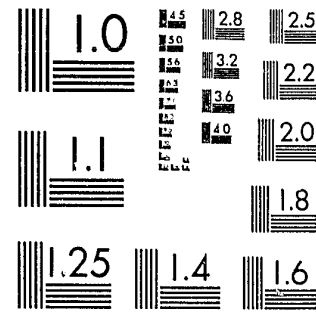


National Criminal Justice Reference Service



This microfiche was produced from documents received for inclusion in the NCJRS data base. Since NCJRS cannot exercise control over the physical condition of the documents submitted, the individual frame quality will vary. The resolution chart on this frame may be used to evaluate the document quality.



MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

Microfilming procedures used to create this fiche comply with the standards set forth in 41CFR 101-11.504.

Points of view or opinions stated in this document are those of the author(s) and do not represent the official position or policies of the U. S. Department of Justice.

National Institute of Justice  
United States Department of Justice  
Washington, D. C. 20531

2/13/85

National Institute of Justice Final Report:

REDUCING BIASES IN JOINED CRIMINAL OFFENSES

Steven Penrod and Sarah Tanford  
University of Wisconsin Purdue University

This research was supported by National Institute of Justice Grant #80-IJ-CX-0048

U.S. Department of Justice  
National Institute of Justice

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Justice.

Permission to reproduce this copyrighted material has been granted by  
Public Domain/NIJ  
U.S. Department of Justice

to the National Criminal Justice Reference Service (NCJRS).

Further reproduction outside of the NCJRS system requires permission of the copyright owner.

96029

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION	1
Legal Background	1
Prejudicial Joinder: Psychological Explanations for Legal Intuitions	6
Prior Research on Joinder	11
The Present Research	16
CHAPTER 2. STUDIES 1 AND 2	20
Study 1 (The Pretest)	20
Study 2 (The Main Study)	24
CHAPTER 3. STUDY 3 (REPLICATION AND EXTENSION)	50
Objectives	50
Results	52
A Comparison of Undergraduates and Representative Jurors	67
CHAPTER 4. DISCUSSION	70
Theoretical Implications	72
Methodological Implications	75
Applied Implications	76
REFERENCES	81

N CJRS

DEC 5 1984

ACQUISITIONS

LIST OF TABLES AND FIGURES

TABLE

2-1. List of cases used in pretest	21
2-2. Trial scenario judgments (pretest)	23
2-3. Experimental designs	25
2-4. Proportion of guilty verdicts	33
2-5. Proportion of statements in each category (target charge)	34
2-6. Correlations among variables	38
2-7. Regression analysis results--Predicting final votes	39
2-8. Initial by final vote distribution	41
2-9. Initial by final vote interactions on post-deliberation data	41
2-10. Juror certainty in verdict as a function of initial and final vote.	43
2-11. Defendant ratings factor analysis and correlation matrix	45
2-12. Correlations between manipulations, ratings and verdicts	46
2-13. Path analysis--Decomposition of causal effects on verdicts	49
3-1. Experimental design (Study 3)	51
3-2. Proportion of guilty verdicts--Target charge (Study 3)	54
3-3. Proportion of guilty verdicts--Charge in 2nd position	55
3-4. Proportion of guilty verdicts--Charge 3	57
3-5. Defendant ratings factor analysis (Study 3) and and correlation matrix	60
3-6. Mean certainty ratings in undergraduates and jury pool subjects as a function of verdict and instructions	64

FIGURE

1. Proportion of guilty verdicts as a function of the number of joined offenses	14
2. Path model of representative jurors' judgment process	48
3. Path model of undergraduate subjects' judgment process	66

## CHAPTER 1. INTRODUCTION

### LEGAL BACKGROUND

Federal Rules of Criminal Procedure, Rule 8(a) Joinder of Offenses. Two or more offenses may be charged in the same indictment or information in a separate count for each offense if the offenses charged, whether felonies or misdemeanors or both, are of the same or similar character or are based on the same act or transaction or on two or more acts or transactions connected together or constituting part of a common scheme or plan.

Federal Rules of Criminal Procedure, Rule 14. Relief from Prejudicial Joinder. If it appears that a defendant or the government is prejudiced by joinder of offenses . . . in an indictment or information or by such joinder for trial together, the court may order an election or separate trial of counts, grant a severance . . . or provide whatever other relief justice requires.

In layman's terms, Rule 8(a) of Federal Rules of Criminal Procedure states that a single defendant may be tried for more than one related offense in a single trial, even if the offenses occur at different times and places and with different victims. The legal term for trying multiple charges together is "joinder of offenses." Rule 14 states that the courts must at the same time protect the defendant from "prejudice" that may result if multiple charges are joined together, but the rule provides little guidance as to how one goes about (1) determining whether prejudice exists and (2) providing "relief" from prejudice. The legal rules do not provide a clear definition of prejudice, although the nature of prejudice has been addressed by legal commentators. For example, Lempert and Saltzburg (1983) describe prejudice as "harm which results when evidence is inappropriately influential because it appeals to the biases or emotions of the fact finder" (p. 156), and McCormick notes that the problem of prejudice arises from "the danger that the facts offered may unduly arouse the jury's emotions of prejudice, hostility or sympathy" (Cleary, 1972, p. 439).

In practice, the legal solution to prejudicial joinder is in the form of a "severance" of offenses. Prior to the trial of a defendant charged with multiple crimes, the prosecutor may seek a joinder of the offenses, and the defense in turn may file a motion of severance, requesting that the charges be tried separately. If this motion is denied by the trial judge and the defendant is convicted of one or more of the offenses, the convictions may be appealed on the grounds of prejudice resulting from the joinder. In both the initial motion for severance and the appeal following conviction, the courts must decide whether joinder was sufficiently prejudicial to warrant separate trials. Because Rule 14 has never been interpreted by the Supreme Court, there are few authoritative guidelines available to judges who must make such decisions. However, convictions resulting from joined trials are often subject to appeal, and there are a large number of published appellate court opinions available at both Federal and state levels. In examining the reasoning used by judges in these decisions, one gets a flavor of the "intuitive psychology" of the legal profession.

To illustrate the court's psychological reasoning we will examine the

case of the United States v. Foutz (1976) which is the leading Federal case on joinder. In this case, the defendant was convicted of two robberies which occurred several months apart, and successfully appealed the convictions on the grounds of prejudice resulting from joinder. In granting the appeal, the judge recognized three possible sources of prejudice that are possible in a joined trial: (1) jurors may confuse the evidence presented in proof of different charges--we will refer to this as the confusion hypothesis, (2) jurors may accumulate or combine evidence across different charges--the accumulation hypothesis and (3) jurors may infer that the defendant has a "criminal disposition" based on the fact that he is charged with multiple crimes--the criminal inference hypothesis.

In applying these three theories of prejudice to the Foutz case, the court noted that the evidence for the second crime was strong while the evidence from the first was weak, so that a jury judging the first offense alone might well have acquitted the defendant. The court thought the jury had probably found the defendant guilty of the second robbery and then concluded that if he had robbed the bank once, there was a good chance he had robbed it before; in other words, they attributed the robbery to the defendant's criminal nature (a criminal inference). In addition, there may have been a "spillover effect of evidence of one crime implicating guilt in other" (accumulation).

At the end of a joined trial, the jury typically receives a special instruction from the judge in addition to the standard jury instructions, the purpose of which is to alleviate potential prejudice resulting from joinder. The exact form of these instructions varies from state to state, but most instructions address at least a portion of the three legal theories of prejudice. The standard Federal joinder instruction reads as follows:

A separate crime or offense is charged in each count of the indictment. Each charge and the evidence pertaining to it should be considered separately. The fact that you may find the accused guilty or not guilty as to one of the offenses should not control your verdict as to any other offense charged (Devitt & Blackmar, Federal Jury Instructions and Practice, 1977, p. 296).

The Federal instruction essentially instructs jurors not to become confused or to accumulate verdicts across charges, but does not instruct jurors to avoid making inferences about the defendant's disposition. The law presumes that instructions will effectively alleviate prejudice, but appellate judges, acting as "intuitive psychologists" do not always agree. In the case of the United States v. Foutz (1976) the court did not think the instruction was sufficient, and the judge quoted an earlier opinion (Bruton v. U.S., 1968) in support of his decision to grant the appeal:

[W]e cannot presume that the jury adhered to limiting instructions and properly "segregated the evidence into separate intellectual boxes."

In other words, the instruction did not eliminate the possibility of confusion of evidence between charges. As a result of the judge's determination that joinder had been prejudicial in the Foutz case, the convictions were reversed and two new, separate trials for each count of

robbery were ordered.

Given that the law recognizes that joinder can be prejudicial, and often is the subject of appeal, the question arises as to the utility of joining charges at all. The main rationale is that it is expedient and saves time and money (Drew v. United States, 1964). However, as the court in Foutz argued, the savings is actually minimal if the evidence for each offense is entirely separate, so that the only real savings is that of choosing only one jury as opposed to more than one.

Thus, the only real convenience served by permitting joint trial of unrelated offenses against the wishes of the defendant may be the convenience of the prosecution in securing a conviction (U.S. v. Foutz, 1976, p. 738).

Another legal precedent for joining charges is the "simple and distinct" test, which holds that joinder will not be prejudicial if the evidence is simple enough that jurors will not become confused (Drew v. United States, 1964). However, even if the assumption that evidential simplicity does reduce confusion is valid, the test does not protect against the other two types of prejudice, accumulation and criminal inference.

Although there is a reasonably large body of case law on joinder, there is little consensus on the criteria that judges ought to apply in their decisions. The issue has not been carefully researched by legal scholars, and there are only a few published legal articles which address the joinder issue. A brief review of the arguments provided in these articles underscores the somewhat conflicting viewpoints among legal scholars concerning the joinder issue.

Remington and Joseph (1961) described some of the conditions under which joinder is generally regarded as appropriate. If several offenses are committed at the same time and place and either damage several victims or do multiple damage to a single victim, it is appropriate to try them together. Joinder is also called for when several offenses occur at different times but are all part of the same scheme or plan. The difficult issue is several unconnected offenses occurring at different times or places with different victims. Remington and Joseph argued that although joinder may be harmful in some circumstances, it may actually be beneficial in others. It may be harrassing to the defendant to defend himself in a number of separate trials, and this could outweigh any disadvantage resulting from joining charges. For this reason and for the reason of expedience, they suggest that a single proceeding may be to the advantage of both sides.

Others have (1) emphasized the potential prejudice to the defendant and (2) argued for clear rules governing joinder. For example, Holderman (1977) discussed the effects of joinder under Nebraska law, which allows similar offenses to be joined even if they are not part of a common plan. The decision to grant severance is left to the judge, and the defendant must be able to demonstrate actual prejudice in order to overturn a conviction. Rather than place this burden on the defendant, Holderman recommended a more stringent test of prejudice, such as that used in United States v. Foutz (1976), in which the test was whether the evidence purported to be prejudicial would have been admissible if the cases were tried separately.

Like Holderman (1977), Baron (1977) advocated the development of clearer rules to govern joinder decisions. In Tennessee law, the decision to join charges is left to the discretion of the judge, and the test of whether the judge has abused this discretion rests on the element of prejudice. However, as in the Federal rule, prejudice is not defined. It has been characterized in a number of different ways by legal commentators, although none of these definitions are very precise (Cleary, 1972). Baron recommended that multiple charges be joined only if they arise out of a single "criminal episode", and that offenses committed on separate occasions not be joined at all. Baron's recommendation is an even clearer guideline than Holderman's "admissibility test," and it would eliminate the joinder situation that legal scholars consider most problematic (Remington & Joseph, 1961), and the one that is most often subject to appeal.

A similar proposal at the Federal level was offered in the Yale Law Journal (Note, 1964-65). This article listed several traditional tests for assessing prejudice arising from joinder, and pointed out the inadequacies of each. The article essentially challenged the intuitive psychological reasoning used by the courts when they conduct a search for absence of prejudice. First, the article questioned whether it is realistic to expect jurors to heed judges' instructions to the jury to confine their decisions to each offense separately. A second common test is "cure by verdict," which assumes that if the defendant is acquitted on any count, the jury must have kept the charges separate, since it was selective in its verdicts. A related test is "cure by concurrent sentencing", which discounts prejudice if the defendant receives one sentence covering multiple counts. The article noted that what both of these "cures" fail to consider is the possibility that the defendant may have been acquitted on all counts if the offenses were tried separately. The final traditional device is that of "overwhelming evidence of guilt" in the record. In other words, if the jury could have reached the same decision on each of the charges tried by itself, then prejudice is not a problem. Using this test, the appellate court in effect becomes the jury, since justices are making judgments about what the jury would have done in a hypothetical situation. The article concluded by stating that the traditional tests of prejudice are simply not adequate, and that the best solution may be to abolish joinder of charges.

Legal scholars may not agree on the solution, but they all agree that joinder is a problem, and that a clear standard is needed to govern joinder decisions. Confusion about the issue among the legal profession no doubt stems from the fact that the conclusions reached by each legal researcher are based on his or her own subjective interpretation of a diverse collection of case law, which is itself a collection of the intuitions of individual judges. From a scientific viewpoint, such an analysis is clearly not an adequate basis for policy formation, and the issue can best be addressed empirically. Of course, non-empirical theorizing dominates the law, and most legal decisions are made in the absence of scientific evidence. Analogous reasoning can be found in the legal responses to problems related to joinder. Therefore, the issues addressed with respect to joinder have additional significance insofar as they suggest other aspects of the trial that could be empirically investigated.

The type of joinder we are concerned with here is joinder of distinct

offenses occurring at different times and places. As noted earlier, a defendant may also be charged with multiple offenses arising out of a single act. In addition to joinder of offenses, the law allows for joinder of defendants, i.e., trying more than one defendant in a single trial. A somewhat related situation occurs when a defendant is charged with a single serious offense and the jury is allowed to simultaneously consider conviction on several lesser included offenses. For example, in some states a jury may be asked to consider a defendant's state of mind or intentions with respect to a homicide, and choose from among first degree murder, second degree murder, and manslaughter. All of the above multiple charge situations may have related effects on jurors' decision processes, all have been the subject of a certain amount of legal theorizing, and all lend themselves to empirical investigation.

The issue of prejudicial joinder is just one example of the intuitive psychological assumptions found in the Rules of Evidence under the more general classification of "prejudicial evidence," and indeed, judges often refer to the Rules of Evidence in their joinder decisions. The psychological implications of prejudicial evidence are discussed by Penrod and Borgida (1983) who observe that the Rules of Evidence recognize several types of evidence that are potentially prejudicial, e.g., character evidence, evidence of other crimes, prior convictions, and similar happenings. In all instances the issue of admissibility is concerned with the relevance of the evidence to the case at hand, and whether relevance outweighs potential prejudice. Rule 401 of Federal Rules of Evidence defines "relevant evidence" as "evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence," and the amount of relevance ascribed to any given evidence is termed its probative value. Rule 403 of Federal Rules of Evidence provides the classic "balancing test" in the law, which is the fundamental rule used to determine admissibility:

Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice.

Unfortunately, the Rules of Evidence do not define prejudice, nor do they provide guidelines for weighing relevance against prejudice. However, the balancing test has been subject to a good deal of legal scrutiny in evidence handbooks and textbooks, since it is such a basic notion in the Rules of Evidence (Cleary, 1972; Lempert & Saltzburg, 1983; Lilly, 1978).

Two categories of prejudicial evidence are especially pertinent to the joinder issue, since they are susceptible to biases similar to those which might result from multiple charges: (1) other crimes and (2) prior convictions. Rule 404 (b) of Federal Rules of Evidence dictates that evidence that a defendant has been involved in crimes other than the one for which he is on trial is not admissible unless it has some relevance to the case other than indicating a propensity to commit crime. In other words, evidence of other crimes should not be used to indicate that the defendant has a criminal disposition--this would be analogous to making a criminal inference as a result of joinder. In fact, appeals of joined-trial convictions are sometimes decided on the basis of whether the charges can pass the "other crimes" test (e.g., Drew v. United States, 1964). Evidence of other crimes is admissible in order to establish that a crime was committed intentionally rather than

accidentally, or to demonstrate the existence of a motive to commit crime. Evidence from other crimes is admissible if there is a common underlying scheme or plan, often an unusual pattern, between earlier crimes and the crime being charged. In a similar manner, evidence of other crimes is admissible if it establishes the identity of the criminal, for example, a similar or unusual modus operandi among the crimes (Lempert & Saltzburg, 1983).

Another type of prejudicial evidence which is relevant to joinder is evidence of a defendant's prior criminal record. Evidence of a prior criminal record is not admissible except for the purpose of attacking witness credibility during cross-examination, although the law recognizes that the introduction of prior convictions even for the purpose of impeachment may well be prejudicial. In fact, Rule 609 of Federal Rules of Evidence includes a special balancing test for evidence of prior convictions. The conviction is not admissible unless its probative value outweighs its prejudicial value. Therefore, the rule reverses the usual pattern for admission of evidence provided in Rule 403, which states that evidence is admissible unless its prejudicial value substantially outweighs its probative value. This suggests that the law recognizes that evidence of prior convictions can be unduly prejudicial, and so takes steps to protect the defendant against such prejudice. In fact, some state courts only allow admission of prior convictions for perjury or for crimes similar to the one being charged, with the rationale that these are the only convictions of sufficient relevance to the defendant's credibility as to outweigh prejudice to the defendant (Lempert & Saltzburg, 1983).

#### PREJUDICIAL JOINDER: PSYCHOLOGICAL EXPLANATIONS FOR LEGAL INTUITIONS

The three legal theories of prejudice: confusion, accumulation, and criminal inference; can be analyzed in psychological terms drawing upon work in the areas of social and cognitive psychology.

##### Confusion

The legal theory of confusion suggests that memory processes might operate to produce different outcomes in joined and single trials. A strictly cognitive explanation is that confusion is the result of interference effects. Research on interference in long term memory has shown that when subjects learn multiple lists of word pairs, they make intrusions (i.e., confuse words) between lists when recalling the lists (Postman & Underwood, 1973). The same sort of interference might occur when a juror is exposed to a series of joined charges, resulting in confusion of evidence among charges.

A more "social" explanation is suggested by research in the area of social cognition. Recent research on person memory suggests that many of our social perceptions have a more cognitive basis in memory (Crocker, Hannah & Weber, 1983; Hastie, 1981; Hastie, Ostrom, Ebbesen, Wyer, Hamilton, & Carlston, 1980). During the trial, a juror is exposed to a large amount of information which must be encoded into a meaningful representation in memory. This memory structure can be termed a schema (Bartlett, 1932; Hastie, 1981). During a joined trial it may not be possible for jurors to encode information for each charge separately, particularly since the trial itself is structured as a single unit containing evidence for all the charges. Jurors may encode a representation of the trial as a whole, especially if the charges are similar,

since similar charges can be assimilated more easily into a single, coherent schema.

Research indicates that specific events (e.g., items of evidence) are recalled in relation to the overall schema (Hastie, 1980). For example, Hastie and Kumar (1979) found that subjects were more likely to recall information that was schema-congruent than schema-incongruent or irrelevant information. Crocker et al. (1983) found that this was primarily the case for recall of behavior attributed to dispositional causes. The criminal inference hypothesis implies that joinder may lead to dispositional attributions, so this suggests that jurors will recall evidence more accurately if each of the joined charges contains different evidence. Research using recognition tasks has indicated that subjects make intrusions, or false recognitions of items that are consistent with the schema they are using. For example, Sulim and Dooling (1974) found that subjects who read a passage about a famous person (e.g., Helen Keller) made more intrusion errors for a sentence that was related to the theme of the passage ("she was deaf, dumb and blind") than for an unrelated sentence ("she was wild, stubborn and violent"). When the main character was not famous (Carol Harris) this effect was not obtained, presumably because no schema was activated. In a joined trial, confusion among charges could take the form of intrusions, i.e., false recognition of evidence from one charge in judgments of other charges.

#### Accumulation

In a related vein, accumulation of evidence might be explained by literature on impression formation which indicates that the amount of information we have about a person affects our overall impression (Schneider, Hastorf, & Ellsworth, 1979). As discussed earlier, one way to analyze the impression formation process is using an information integration model (Anderson, 1974, 1978) which postulates that the overall impression is a weighted average of individual items of information. Joinder might affect the weight assigned to different kinds of trial information. Jurors judging a joined trial may weigh the evidence for the prosecution more heavily than evidence for the defense, or they could weigh evidence about the defendant's character more heavily than would ordinarily be the case. Joinder could also affect jurors' initial impression of guilt, prior to hearing any trial evidence.

There is evidence to indicate that negative information is weighted more heavily than positive information (Anderson, 1965; Fiske, 1980; Hamilton & Huffman, 1971; Hodges, 1974; Kanouse & Hanson, 1972). This implies that negative trial evidence may accumulate at a faster rate than positive evidence. Since jurors in a joined trial are exposed to more evidence than jurors judging a single charge, their impression of the defendant at the end of the trial should be less favorable.

It is possible that joinder creates context effects in the impression formation process akin to the context effects initially investigated by Asch (1946) using sets of trait adjectives. A number of researchers have investigated context effects within an information integration framework (Anderson, 1966; Anderson & Lampel, 1965; Kaplan, 1971, 1974, 1975; Ostrom, 1977; Wyer, 1974). Anderson (1966; Anderson & Lampel, 1965) gave subjects sets of three trait adjectives, consisting of one test adjective and two adjectives

which served as the context (one of four levels of favorability). Subjects were asked to rate the favorability of the test adjective and their overall impression of the person described by the adjectives. Anderson found that both trait and impression ratings were affected by the valence of the test adjective and by the context, and that these two effects were additive and therefore consistent with a weighted averaging model. The context did not change the meaning (scale value) of the test item, instead it was averaged in with the value of the test item to form the impression. A similar process could operate in a joined trial in which the context created by multiple charges results in an unfavorable overall impression, which is then averaged in with evaluations of the evidence (and perhaps weighted heavily).

Let us consider an alternative explanation for the accumulation of evidence process. In line with the above reasoning, we assume that the multiple charge situation creates an unfavorable context within which subsequent information is interpreted. However, this time we do not assume that jurors average together items of evidence in a rational manner. Instead, the juror is viewed as a cognitive miser (Taylor & Crocker, 1980) who selectively processes information using various shorthand devices (Kahneman et al., 1982; Nisbett & Ross, 1980). Hamilton (1981) argued that the emphasis on quantitative models of information integration has obscured our understanding of the underlying processes involved in impression formation. Hamilton proposed that the cognitive representation of a person can be considered a social schema which guides the encoding, organization and interpretation of information about the person (Taylor & Crocker, 1980). In a detailed review of research on social schemas, Taylor and Crocker (1980) cite research which indicates that people accept data as being consistent with their schemas even when it is neutral or inconsistent. People distort information to make it consistent with the schema (Langer & Abelson, 1974; Zandney & Gerard, 1974) and resist information that disconfirms the schema (Lord, Ross, & Lepper, 1979; Ross, Lepper & Hubbard, 1975). To apply a schematic impression formation analysis the joinder situation, we assume that the negative context created by the multiple-offense trial results in an unfavorable impression of the defendant which could be termed a "criminal schema." This schema will affect the interpretation of incoming evidence, causing jurors to judge the evidence against the defendant as stronger and disregard evidence that indicates innocence (and is thereby inconsistent with the criminal schema). The result of this biased information processing will be an accumulation of evidence that indicates the defendant's guilt.

#### Criminal Inference

Attribution theory (Kelley, 1967, 1972) provides a possible explanatory model for the legal theory of criminal inference. Kelley's (1967) covariation model is a particularly useful framework for analyzing jurors' judgments of a defendant charged with multiple offenses, since it is concerned with inferences made about the causes of behavior for which there are multiple observations. First, the behavior of a defendant charged with multiple crimes can be viewed as low in distinctiveness--the person is charged with criminal acts with respect to several different entities (victims, situations, times, places). Second, the defendant's behavior can be considered high in consistency--the person has allegedly displayed criminal behavior on more than one occasion. Finally, criminal behavior in all instances can be considered low in consensus--most people are not criminals. These three

components; low distinctiveness, high consistency, and low consensus; are the requirements for an internal attribution in Kelley's model. In contrast, when judging a defendant charged with a single crime, jurors do not have distinctiveness and consistency information. Kelley's (1972) causal schemata model was designed for situations in which there is only information about a single instance of behavior. In the single-charge situation jurors may apply the discounting principle (Kelley, 1972), whereby people discount possible causes for an event to the extent that there are other, more plausible causes available. There are more alternative plausible explanations for being charged with a single crime than with several. For example, it is more likely that a suspect apprehended near the scene of a single crime just "happened to be at the wrong place at the wrong time" than a suspect apprehended near the scenes of several crimes.

Although multiple-charge judgments can be analyzed rather handily using Kelley's attribution model, the theory has a number of limitations which suggest that it may not be the best explanation for inferences made about the defendant. One criticism is that people frequently underutilize consensus information when making causal inferences (Kassin, 1979; Nisbett, Borgida, Crandall, & Reed, 1976). In addition, research has indicated that people process information selectively, and that attributions are influenced by the most salient features in the situation (Taylor & Fiske, 1978). A simple change in focus of attention can be sufficient to change an individual's attributions (Taylor & Fiske, 1975). Another question is whether people are even able to accurately assess covariation if given sufficient information. A review of research by Crocker (1981) indicates that the layperson has rather limited ability to assess covariation. The same point is made by Nisbett and Ross (1980), who devote an entire volume to an analysis of limitations in the inference process. According to Nisbett and Ross, individuals' causal judgments are "theory driven" rather than "data driven", and are thereby subject to systematic biases. Biases in the attribution process often result from the use of heuristic devices (Ross, 1977; Tversky & Kahneman, 1974).

It was noted previously that heuristics are likely to play a part in jurors' decisions, since the trial is a situation of decision making under uncertainty. Here we consider specifically how heuristics could operate to bias jurors' judgments of a defendant in a joined trial.

(1) Representativeness. In discussing the impression formation process it was hypothesized that joined trials activate a "criminal schema," i.e., an impression of the defendant as a prototypical criminal. To the extent that jurors base their judgments on representativeness, they will be more likely to form a judgment of guilt in a joined trial, since the defendant is considered representative of criminals in general. (2) Anchoring and adjustment could operate in two different ways. First, people tend to overestimate the probability of compound events (Bar-Hillel, 1973) and may overestimate the likelihood of guilt on multiple counts, due to insufficient adjustment. Second, people tend to anchor their judgments in the direction of the initial value (Tversky & Kahneman, 1974). If the prosecutor is a good social psychologist and tries his strongest case first, this might produce an anchoring effect in the direction of guilt which will affect judgments of subsequent charges. (3) Availability might operate by making especially salient the criminal characteristics of the defendant based on the fact that he is charged with multiple crimes. Jurors may then focus on these

nonevidentiary factors rather than the evidence, and make causal inferences based on the defendant's criminal disposition. If, as we proposed earlier, joinder activates a criminal schema, this will affect memory for the evidence by making unfavorable information more available. Availability could also operate if multiple charges consist of one strong case and other weaker ones (as is often the case). The stronger evidence may be more memorable, and hence could affect judgments to a greater degree than weaker evidence.

#### Judges' Instructions

The judge in a joined trial instructs jurors to reach a decision for each charge independently, and the legal intuition is that they will be able to do so. This is one particularly good example of a situation for which social psychological research indicates that legal intuitions are probably incorrect. For example, Ross's work on belief perseverance has indicated that people often have difficulty disregarding information even after it has been totally discredited (Lord, Ross, & Lepper, 1979; Ross, Lepper, & Hubbard, 1975). This is reminiscent of earlier work on the sleeper effect (Hovland, Janis, & Kelley, 1953) which indicated that people may not be affected by discredited information initially, but are influenced later after the source is forgotten. However, in judgments of joined trials we contend that the negative impression is formed at the time the information is received, due to context effects in impression formation (Asch, 1946) or to the activation of a criminal schema. Judges' instructions come at the end of the trial, and it is difficult to imagine that they will do anything to change the impression already formed. In fact, instructions might even increase prejudice by making the prejudicial information more salient and hence available in memory. It is also unlikely that instructions presented at the end of the trial will prevent confusion of evidence that is presented during the trial.

With respect to accumulation of evidence, research on primacy effects in impression formation (Anderson, 1965; Luchins, 1957) suggests that early information is often weighted disproportionately to later information. Therefore, the accumulation process may begin early in the trial, suggesting that instructions will not effectively reduce accumulation. Kassin and Wrightsman (1979) obtained support for a primacy effect in a jury simulation study which demonstrated that reasonable doubt instructions influenced jurors' judgments when presented before, but not after the trial.

#### Summary and Integration of Approaches

In the preceding pages a number of cognitive and social psychological approaches have been used to analyze each of the three legal theories of prejudice resulting from joinder, as well as the legal remedy of judges' instructions. We discussed ways in which joinder will affect memory, impression formation processes, and causal inferences. For each of the three processes, we first discussed how it might operate if decisions are made in a rational, scientific manner. The same process was then analyzed in terms of recent theorizing in social cognition which indicates that social information processing may not be so rational (Higgins, Herman, & Zanna, 1981; Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980).

### Integration of Processes

Social cognition research suggests that the legal notions of confusion, accumulation, and criminal inference may be recast into an explanatory social psychological model. We postulate that the multiple-charge situation produces an unfavorable impression of the defendant which influences memory for the trial by (a) promoting confusion among charges, and (b) making unfavorable information more available in memory. The initial impression leads to an accumulation of evidence because jurors (a) distort evidence to make it consistent with the impression, (b) ignore evidence that contradicts the impression and (c) make insufficient adjustments when combining probabilities based on multiple offenses. Finally, the unfavorable impression is both the product the source of inferences made about the defendant. Jurors initially form an impression that the defendant has a criminal disposition based on the fact that he is charged with multiple crimes. The criminal impression then affects inferences of the likelihood of guilt based on judgments of representativeness. The impression of guilt that results from this process is quite resistant to change in response to judge's instructions. Research by Smith and Miller (1983) provides empirical evidence the inferences about an actor's disposition occur prior to judgments about the causes of an event, and may therefore mediate these judgments. Smith and Miller's results support a schema-based model of causal inference.

### PRIOR RESEARCH ON JOINDER

In the first section of this paper the legal theories pertaining to joinder and other types of prejudicial evidence were outlined. Then a number of psychological mechanisms were proposed to account for biases in jurors' judgments. In this section empirical research investigating the effects of joinder is reviewed. Some of the research has been rather atheoretical, with the goal of demonstrating that a particular type of evidence has an effect, rather than explaining what mediates the effect. Where possible, the results of these studies are interpreted in terms of the theoretical perspectives discussed above.

Horowitz, Bordens and Feldman (1980) examined the impact of joining two charges on jurors' assessments of defendant guilt using an audiotaped trial summary. Horowitz et al. employed a 2 (strength of evidence) x 3 (severed or joined with a strong or weak case) x 2 (position) factorial design. All main effects were significant as were several interactions. The significant joinder effect demonstrated that jurors' ratings of defendant guilt were higher when two offenses were joined than when the offenses were tried separately. On one of the two rape charges examined in the study subjects rated the likelihood of guilt as .39 when tried as a separate charge and .46 for the same charge when it was joined with another. On the other offense they rated guilt as .41 when tried separately and .48 when joined. Horowitz et al. reported that this effect occurred primarily for offenses presented first rather than second, and suggested that this may arise because jurors suspend judgment on the first case until they have received evidence on the second offense, which serves as a kind of anchor. However, in examining their results it appears that in some conditions the joinder effect was also operating in the second case. In conditions in which the evidence was clear (as opposed to close), subjects rated guilt on one offense as .44 when severed and .49 when joined (a significant difference), and on the other offense they

rated guilt as .48 when severed and .53 when joined (not significant but in the predicted direction). Also, the severed case was always judged individually, and so technically could not appear in the second position, but was treated as such in order to complete the factorial design. Therefore, comparisons between the second severed and joined offenses could not meaningfully be made. Even so, the joinder effects obtained were quite strong --our estimates of effect size (Epsilon--Cohen, 1977) for their two cases were .36 and .35 overall. Horowitz et al. offered no evidence as to the process by which prejudice might operate in their joined cases.

Bordens and Horowitz (1983) investigated the effects of joining two charges that varied according to case strength, charge similarity (two rapes or a murder and a rape) and case order, again using audiotaped trial summaries. They found that convictions on the first but not the second charge were significantly higher when the charge was joined than when it was severed. Convictions were also more likely when the joined charges were similar than when they were dissimilar. Bordens and Horowitz investigated the processing of trial information by asking subjects to recall evidence from the cases, generate thoughts that related to their verdict preference, and rate the thoughts according to their degree of favorableness to prosecution or defense. Although the reported results are complex, a general pattern was detected. Subjects generated a greater number of thoughts and, more importantly, a higher percentage of anti-defendant thoughts when cases were similar than when they were dissimilar. Subjects also made a higher percentage of anti-defendant recall intrusions from the second case to the first when charges were similar. The charge similarity effects for percentages of both antidefendant intrusions and thoughts were obtained primarily when the second case was "close" (ambiguous in strength) rather than "clear" (strongly in favor of the prosecution).

Interestingly, Bordens and Horowitz found that ratings of thoughts against the defendant did not differ in joined and severed conditions. These ratings can be considered an indirect measure of the strength of evidence against the defendant, suggesting that joinder did not affect judgments of evidence strength. However, ratings of thoughts against the defendant bore a strong relationship to verdict ratings ( $r = .76$ ), suggesting that all subjects' judgments were strongly influenced by their assessments of evidence strength. The proportion of anti-defendant thoughts were also positively related to verdicts, as were the proportion of intrusions against the defendant ( $r$ 's not provided). This result is interesting in light of research reported below (Tanford & Penrod, 1982), in which intrusions bore little relationship to guilt judgments.

In two experiments conducted by Greene and Loftus (1981) subjects read excerpts from a trial transcript consisting of a single charge (murder or rape) or two charges (murder and rape). Greene and Loftus found that the defendant was more likely to be convicted of either crime if the two charges were joined than if they were tried separately. Judges' instructions to consider charges separately were ineffective in removing this effect, regardless of whether they came before or after the trial. Greene and Loftus investigated three mechanisms to account for their effects: memory, a change in reasonable doubt standard, and inferences of a criminal disposition. They found that subjects in joined and single conditions were equally accurate on a fact recognition task, so the memory explanation was not supported. There was



also no difference between joined and single conditions in the amount of proof needed to convict the defendant. However, subjects in joined conditions rated the defendant more negatively on the dimensions of dangerousness, likeableness, and believability, suggesting that joinder affected inferences about the defendant's character.

Kerr and Sawyers (1979) examined the independence of judgments of multiple charges, particularly the way the strength of evidence on one charge affects the judgment of the charge with which it is joined. They found that as the strength of evidence on one charge increased, the probability of conviction for the other charge tended to decrease. They argue that this does not support an accumulation of evidence model but does support an equity model, in which jurors want to produce a pair of verdicts that is fair. However, a point which was not central to their study is central to the present concerns. As Kerr and Sawyers (1979) noted in their discussion, a variant of the accumulation of evidence hypothesis, namely an "accumulation of charges" was supported. Conviction on the weak charge alone was determined to be about 25% through pretesting. The conviction rate was much higher in the experimental conditions, all of which consisted of two charges. The mean conviction rate for the weak robbery charge when joined was 57%. A joinder effect was not obtained for strong charges (although the possibility of a ceiling effect cannot be ruled out). A possible explanation for the lack of joinder effects with strong cases is that prejudicial factors may be more salient, and hence available, when other evidence is weak.

Bordens and Horowitz (1983), Greene and Loftus (1981), Horowitz et al. (1980) and Kerr and Sawyers (1979) all found that a defendant was more likely to be convicted in a trial of two joined charges than when tried for a single charge. In a study designed to investigate each of the three legal theories of prejudice, Tanford and Penrod (1982) extended this finding to trials with three and four offenses. Subjects read written trial summaries consisting of a single charge or a joined trial of two, three or four charges in one of several combinations. The results indicated that the probability of conviction on a particular charge increased as a function of the number of charges with which it was joined. These results are illustrated in Figure 1 (from Tanford & Penrod, 1982). Tanford and Penrod also obtained support for each of the three legal theories of prejudice resulting from joinder. Subjects in joined trials made significantly more intrusion errors in recall of evidence from a joined trial of three charges than in recall of evidence from three separate charges containing the same information. In other words, subjects tended to confuse evidence from different charges in the joined trial. Subjects judging joined offenses also rated individual items of evidence as more incriminating than subjects who rated the same evidence from a single-offense trial, supporting the process of accumulation of evidence. The mean rating of evidence strength was 4.63 in joined conditions and 3.95 in single conditions (9-point scale). Subjects also rated the defendant on seven 9-point bipolar scales on a number of trait and behavioral characteristics. On all ratings the defendant was rated less favorably in joined than single conditions, and 5 out of 7 of these differences were statistically significant. This supports the theory of criminal inference.

Tanford and Penrod also examined the relationship between memory,

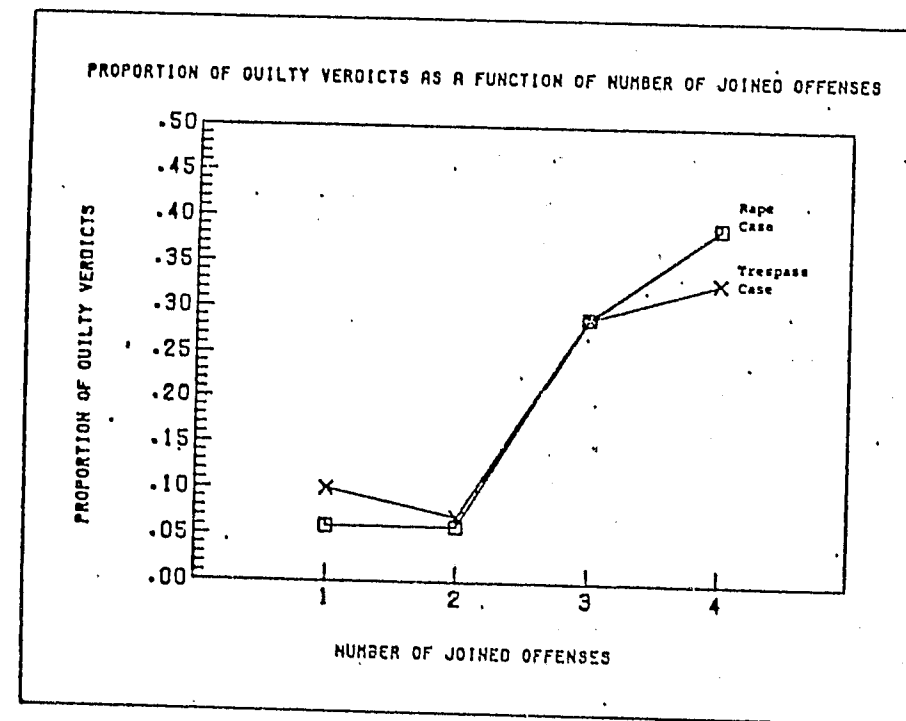


Fig. 1. Proportion of guilty verdicts as a function of the number of joined offenses.

evidence strength and defendant ratings with respect to verdicts and judgments of the defendant's guilt. They found a low, nonsignificant positive relationship between memory intrusions and guilt, suggesting that confusion is not a key mediating factor. They found a strong positive relationship between ratings of the evidence and guilt in both joined and single conditions--the mean correlation was .52 for single charges and .51 for joined charges. Finally, they found a positive relationship between defendant ratings and guilt that was stronger for joined charges (mean  $r = .54$ ) than for single charges (mean  $r = .33$ ). What these results suggest is that all subjects were basing their judgments on the strength of the evidence, as they are supposed to legally (although the results were correlational, so the direction of these effects is not known). In addition, subjects based their judgments on inferences about the defendant, and subjects in joined trials did so to a greater degree than subjects judging single trials.

Tanford and Penrod's results can be interpreted in terms of the schematic processing of trial information. The defendant ratings suggest that joinder activated a criminal schema which affected the processing of trial information (as indicated by recall and evidence ratings) and which, along with the evidence, influenced guilt judgments. The correlations suggest that these judgments may have been due, in part, to the fact that the defendant in the joined trial was considered representative of the prototypical criminal, whereas in the single trial he was not.

#### Limitations of Previous Research

The research reviewed on the effects of joinder (and related prejudicial evidence) demonstrates empirically that mock jurors' judgments can be biased by several evidentiary and procedural factors. The law has posed a number of questions in the Rules of Evidence and Rules of Criminal Procedure concerning prejudicial evidence and trial procedures, and has suggested remedies for alleviating prejudice. Jurors are to be instructed to disregard prejudicial evidence, and judges are told to weigh relevance against prejudice in determining admissibility. The empirical studies have essentially confirmed legal intuitions that evidence and procedures can be prejudicial, and have further suggested that the legal remedies may not be adequate. However, these studies suffer from a number of limitations, both in terms of applications to the courtroom and in terms of providing an understanding of the psychological mechanisms underlying judgment biases.

From an applied standpoint, most of the studies reviewed were conducted using procedures which were low in external validity, so that their generalizability to actual trial settings is questionable. The methodological issues involved in conducting jury simulation research are discussed at length elsewhere (Bray & Kerr, 1982; Special Issue: Simulation research and the law, 1979), and are not detailed here. We merely point out the limitations of the empirical studies which may preclude the applicability of their findings to the courtroom. First, most of the studies reviewed used written trial summaries as stimulus materials. Therefore, the trial presentation for the majority of studies was not representative of an actual jury trial. Second, most of the studies used college students as subjects. Therefore, none of the studies employed subjects who were truly representative of the juror population. Finally, few of the studies included group deliberation in their procedures. For studies which specifically examined joinder effects, all used

undergraduate subjects, all used written trial summaries except for the Bordens and Horowitz (1983) and Horowitz et al. (1980) studies which used audiotaped summaries, and none included group deliberation. In order to obtain results that can be applied to the courtroom, it would seem desirable to conduct experiments more closely resembling an actual trial.

The limitations of previous research are at the same time theoretical ones. The research may provide information about judgments made by undergraduates responding to a questionnaire, but it does not provide substantial insight into the psychological processes in operation for jurors within the social context of an actual jury trial. The research is also limited in that it has generally not been theoretically motivated, instead the goal has been to demonstrate an effect for some type of evidence or procedure. We have suggested interpretations for some of the results, but clearly research is needed to investigate the factors mediating these effects.

Studies on joinder effects have made some effort to investigate the processes involved in multiple-charge judgments. Tanford and Penrod (1982) and Bordens and Horowitz (1983) found that joinder affected memory for the trial, and Bordens and Horowitz found that memory was related to guilt judgments, whereas Tanford and Penrod found that it was not. Greene and Loftus (1981) and Tanford and Penrod (1982) found that joinder led to negative inferences about the defendant, and Tanford and Penrod found that these inferences were related to guilt judgments, particularly in joined trial conditions. Tanford and Penrod also found that joinder affected judgments of the strength of the evidence, which in turn were related to guilt judgments in both joined and single trials. Of course, all studies on joinder found that joinder increased the probability of conviction, but Bordens and Horowitz (1983) and Horowitz et al. (1980) obtained this effect primarily on the first joined charge, whereas Greene and Loftus (1981) and Tanford and Penrod (1982) obtained increased conviction rates on all joined charges, regardless of position. Therefore, on the basis of existing research, there is ample evidence that joinder does bias judgments, and there is some evidence as to the processes involved.

#### THE PRESENT RESEARCH

##### Overview of the Present Research

The present research was designed to avoid the limitations of previous research by using procedures that maximized external validity, and by investigating the underlying mechanisms involved in jurors' judgments. The purpose of the research was to examine the effects of multiple charges using a realistic trial simulation in order to obtain results that could be applied to the courtroom, and to provide an understanding of the psychological processes in operation for actual jurors judging a joined trial. The research was designed with several general goals in mind. One goal was to determine whether the results obtained in earlier laboratory experiments could be replicated and extended in a more realistic trial setting. A second goal was to develop a research paradigm that could later be used to investigate other assumptions in the Rules of Evidence, using methods and procedures that would have clear applicability to the courtroom. As a corollary to this, it was intended that the results obtained would provide insight into the general processes involved in judgmental biases in the courtroom.

Specifically, the purpose of the research was to determine under what conditions (and to what extent) jurors will become biased when several charges are tried together in a single trial. It was predicted that joinder would increase the likelihood of conviction, but that the magnitude of these effects would be influenced by three independent variables: (1) The similarity of the offenses charged, (2) the similarity of the evidence contained in the offenses and (3) judges instructions designed to reduce prejudicial effects of joinder. Similarity was examined for two reasons. First, from an applied perspective, we wanted to provide guidance to judges as a basis for making decisions about when to join charges. The law primarily allows for joinder of similar charges, though in fact joining similar charges might be more prejudicial than joinder of dissimilar charges. Because the courts have looked to similarity as a basis for categorizing charges, the present research was designed to inform the courts about what specific combinations of charge and evidence similarity would be most prejudicial. Second, social psychological research as well as previous research on joinder suggested that similarity would affect the relative contribution of the hypothesized sources of prejudice: confusion, accumulation and criminal inference.

Hypotheses. Based on research and theory in social and cognitive psychology, as well as empirical research on joinder and other prejudicial evidence, a number of predictions were made concerning the effects of multiple charges.

1. It was predicted that (a) a defendant charged with three offenses was more likely to be convicted on any particular charge than a defendant tried for the same crime by itself. Contrary to the findings of Bordens and Horowitz (1983) and Horowitz et al. (1980) this effect was (b) predicted not to be specific to the first joined charge. Instead, joinder was predicted to increase convictions on all charges. (c) It was also predicted that subjects in joined trials would confuse evidence among charges, view the evidence as stronger than subjects in single trials, and make negative inferences about the defendant.

2. The similarity of the joined offenses was predicted to influence the magnitude of the conviction effects by influencing the memory and social inference processes hypothesized to mediate joinder effects. Bordens and Horowitz (1983) found that a defendant was more likely to be convicted in a trial of two rape charges than in a trial of rape and murder, and that subjects made more recall intrusions when charges were similar. In line with these results, it was predicted that (a) a defendant charged with three similar crimes would be more likely to be convicted than a defendant charged with three dissimilar crimes. (b) Greater confusion between charges was predicted when charges were similar. (c) It was also predicted that subjects would make more inferences about the defendant's criminal character when charges were similar than when charges were dissimilar. A series of similar charges is more likely to evoke a criminal schema. In attribution theory terms, being charged with several similar crimes creates an impression of consistency.

3. The similarity of the evidence contained in the joined offenses was predicted to affect jurors' judgments. (a) Evidence similarity should primarily affect the accumulation of evidence process, since there will be more evidence to accumulate if the evidence for each charge is different. (b)

However, more confusion of evidence was predicted for charges containing similar evidence. (c) Although it is not clear that evidence similarity will affect inferences about the defendant, attribution theory suggests that more dispositional attributions might be made when evidence is dissimilar, indicating behavior that is low in distinctiveness. (d) It was predicted that similar and dissimilar evidence might also vary in terms of their probative value, credibility, or informativeness, and these dimensions could also affect jurors' judgments.

4. It was predicted that the three processes of confusion, accumulation and criminal inference would be related to subjects' guilt judgments. Previous research (Tanford & Penrod, 1982) indicated that confusion was not related to verdicts, although Bordens and Horowitz (1980) found that it was. Tanford and Penrod (1982) also found that judgments of the defendant and evidence were related to judgments of guilt. The present research was designed to assess the relative contribution of each of the three processes to the prejudicial effects of joinder.

5. Another research question was whether judges' instructions could effectively reduce prejudice resulting from joinder. Research in social psychology on belief perseverance (Lord, Ross, & Lepper, 1979; Ross, Lepper, & Hubbard, 1975), context effects (Asch, 1946) and the resilience of schemas (Taylor & Crocker, 1981) suggests that instructions will probably not be effective, as does the bulk of empirical research on judges' instructions (Lind, 1982). Nevertheless, we did want to give the legal remedy for prejudice a fair test, so we designed a very strong and complete set of instructions in order to make a more conclusive statement about their effectiveness than those made in previous studies.

6. It was predicted that group deliberation would affect jurors' decisions. Although there is some research to suggest that deliberation can correct juror bias (Kaplan & Miller, 1978), research on group polarization suggests that deliberation might serve to aggravate joinder-induced biases (Lamm & Myers, 1978). In addition to affecting judgments of guilt, it was predicted that deliberation would affect memory, impression formation and social inference processes.

7. A somewhat subsidiary question was whether representative juror subjects and undergraduates differ in their judgement processes. It was predicted that both subject populations would be affected by joinder, but that the magnitude of these effects might differ (Linz et al., 1981). The two groups might also vary in their susceptibility to similarity and instruction manipulations.

Plan of the Research. Three studies were designed to test these hypotheses. Study 1 was essentially a pretest for the stimulus materials used in Studies 2 and 3. Undergraduate subjects read several case summaries, made guilt and verdict judgments, and rated the similarity of charges and evidence and the strength of the evidence. Study 1 also allowed a comparison of these ratings with varying amounts of trial context provided.

Study 2 can be considered the "main study" of the research. Qualified juror subjects viewed realistic videotaped trials presented either as a single offense or in a joined trial of three charges that varied according to charge

similarity, evidence similarity and judges' instructions. Subjects gave an individual verdict preference, deliberated in groups of six and reached a group decision, and responded to a questionnaire designed to assess memory, accumulation of evidence, and criminal inferences. Additional subjects completed the same questionnaire without deliberation.

Study 3 replicated the experimental conditions from Study 2 using undergraduate subjects who did not deliberate. The study included several additional experimental conditions that could not be included in the main study due to time and financial constraints. Finally, the data from Studies 2 and 3 were combined in a single analysis in order to compare undergraduate and qualified juror responses.

## CHAPTER 2. STUDIES 1 AND 2

### STUDY 1 (THE PRETEST)

#### Method

The primary objective of the first study was to assure that the stimulus materials for Studies 2 and 3 met two criteria. First, the individual offenses should be on the weak side, to preclude the possibility of ceiling effects when the charges were combined into a joined trial. Second, the offenses should vary as a function of charge and evidence similarity, so that similarity could be manipulated when offenses were joined together.

A secondary goal was to investigate subjects' ratings of trial evidence under varying levels of context information. A weighted averaging model of information integration (Anderson, 1974) predicts that context and evidence strength should have additive effects. Since the cases were designed to be weak, this suggests that evidence judged in the context of the trial should be rated as less incriminating than evidence judged by itself.

Stimulus materials. Thirteen written trial scenarios consisting of nine burglaries, two armed robberies and two assaults and batteries were adapted from reports of cases tried in Wisconsin. Common, everyday crimes were used in order to enhance the generalizability of the findings. The particular offenses were chosen after consultation with representatives from the State Public Defender's office, the Attorney General's office, and the University Legal Defense Project, all of whom indicated that burglaries, robberies and assaults were among the cases most frequently heard by juries in Wisconsin. Each scenario included a brief definition of the charge, opening statements by prosecution and defense, direct and cross-examination of witnesses, closing arguments by the attorneys, and judges' instructions on the law. The scenarios were four to seven pages in length, and averaged between 1200 and 1800 words. For each case, a short summary version was also prepared. The summary consisted of a two to three sentence fact summary describing the general nature of the crime, followed by one to four sentences describing specific pieces of evidence for prosecution and defense. The crimes were designed to vary according to the type of crime charged and the type of evidence presented, so that they could be joined together in different combinations of charge and evidence similarity for the main study. For purposes of pretesting they were treated as completely separate charges with different defendants. Table 2-1 presents a summary description of the cases used, grouped according to charge and evidence similarity.

Procedure and subjects. 82 undergraduates at the University of Wisconsin participated in the study for course credit. Subjects read and judged the trial materials in one of three conditions: (1) trial scenario. Forty-one subjects read complete trial scenarios for seven of the offenses randomly selected from the total of 13 cases, presented in a random order. Due to the length of the materials, it was not feasible to ask subjects to read all 13 cases. Each particular offense was judged by at least 19 subjects. (2) case summary. Twenty subjects read the brief summary versions for each of the 13 cases, presented in a random order. (3) evidence only. Twenty-one subjects read and rated individual items of evidence from the cases (53 items in all) arranged in a random order with no trial context provided.

Table 2-1. List of cases used in pretest

Charge #	Crime charged	main evidence	Charge #	Crime charged	main evidence
0	Burglary -- service station (target charge)	circum- stantial			
1	Burglary -- service station	circum- stantial	7	Burglary -- service station	Circum- stantial identical charges similar evidence
2	Burglary -- service station	eyewitness I.D.	8	Burglary -- service station	Possession of stolen money identical charges dissimilar evidence
3	Burglary -- mobile home	circum- stantial	9	Burglary -- pharmacy	circum- stantial similar charges similar evidence
4	Burglary -- warehouse	eyewitness I.D.	10	Burglary -- home	informant similar charges dissimilar evidence
5	Assault and battery	circum- stantial	11	Armed robbery	circum- stantial dissimilar charges similar evidence
6	Assault and battery	eyewitness I.D.	12	Armed robbery -- possession of stolen goods	dissimilar charges dissimilar evidence

Dependent variables. After reading each case, and before going on to the next, subjects were asked to make the following judgments about the offense: (1) verdict preference (guilty or not guilty), (2) probability of the defendant's guilt on a 9 point scale (1 = definitely not guilty, 9 = definitely guilty), (3) strength of the evidence overall on a 9-point scale (1 = extremely weak, 9 = extremely strong) and (4) the incriminating value of several individual items of evidence on a 9-point scale (1 = strongly indicates innocence, 9 = strongly indicates guilt). After reading and judging all the cases, subjects rated all possible pairs of cases on three dimensions of similarity: (1) overall similarity, (2) charge similarity and (3) evidence similarity. These ratings were made on a 9-point scale where 1 = very similar and 9 = very dissimilar. Subjects in the evidence only condition rated the same individual evidence items as subjects in the other two conditions, and did not provide any other judgments.

#### Results

In terms of pretesting the stimulus materials for use in the main study, the primary concern was with judgments of the complete trial scenario. Table 2-2 presents the proportion of guilty verdicts, mean probability of guilt ratings, and ratings of overall evidence strength for prosecution and defense, for each of the offenses. Our goal was to devise cases that were on the weak side while still allowing for variability in subjects' responses. The mean proportion of guilty verdicts was .26, with a range from 0 to 47%, so we were successful in selecting cases that met this criterion. The proportion of guilty verdicts for the case designated the "target offense" (case 0) was .31. The probability of guilt ratings, which ranged from 3.19 to 5.79 on a 9-point scale with a mean of 4.74 indicated that the cases were fairly ambiguous as to guilt. This is further corroborated by the ratings of evidence strength overall, with a mean rating of 4.86 for prosecution evidence and 5.32 for defense evidence.

Prior to pretesting the cases were classified according to the charge and evidence similarity levels to be used in Studies 2 and 3. These are the groupings in Table 2-1, along with case 0, the target offense, which was to be included in each combination. Of course, these charge combinations were not grouped together for pretesting--the offenses were presented as individual trials ordered randomly. In order to assess the accuracy of our operational definitions of similarity, mean similarity ratings were computed for the three levels of charge similarity and two levels of evidence similarity. The mean charge similarity ratings were 4.25 for highly similar ("identical") charges, 5.06 for moderately similar charges and 6.21 for dissimilar charges, where a lower number indicates greater similarity. The mean evidence similarity rating was 4.92 for charges with similar evidence and 5.58 for dissimilar evidence. Therefore, the classification of the offenses in terms of both charge and evidence similarity was as predicted. However, it was noted earlier that levels of evidence similarity might also differ along other dimensions, for example, informativeness or credibility. The evidence from different offenses was not rated on dimensions other than similarity at the pretest stage. Supplementary ratings will be presented later in this paper to shed some light on this question.

Table 2-2. Trial scenario judgments (pretest)

Case	Proportion of guilty verdicts	Probability of guilt	Prosecution evidence	Defense evidence	N
0 (target)	.31	5.46	5.49	5.28	39
1	.19	3.96	4.05	4.81	21
2	.45	5.55	5.50	3.90	20
3	.47	5.79	5.74	4.58	19
4	.47	5.62	5.90	5.19	21
5	.19	4.33	4.29	5.71	21
6	.20	3.70	3.95	5.70	20
7	.00	3.19	2.81	6.67	21
8	.14	3.86	3.86	5.43	21
9	.05	4.45	4.64	5.59	22
10	.05	4.52	5.38	6.14	21
11	.40	5.25	5.50	5.65	20
12	.42	5.58	5.68	4.32	19

STUDY 2: THE MAIN STUDY

Method

Subjects. A total of 732 subjects participated in the experiment. Of these, 714 were qualified jurors who had been summoned for service in the Dane County, Wisconsin, jury pool for 1981 and 1982. Of these, 492 (69%) subjects had jury experience, while the remaining 31% had been summoned but not seated on a trial. Jurors were first sent a letter describing the study and were followed up with a phone call to schedule them for a session. Jurors were paid \$20 for participation. The remaining 18 subjects were undergraduates at the University of Wisconsin who were registered voters and therefore jury-qualified. Undergraduate subjects received Introductory Psychology course credit for participation, with the exception of one subject who received \$20. Undergraduates were scheduled in order to fill in sessions for which there were not enough jury pool subjects to form six-person groups for deliberation. Twelve groups contained a single undergraduate and three groups contained two undergraduates, and these groups were distributed evenly across the experimental conditions. The sample as a whole was 49% female and 51% male. The mean age was 40 years, with a range from 18 to 82. Subjects represented wide range of socioeconomic status variables such as income, occupation and educational background.

Design. The design of the experiment is presented in Table 2-3. A partial factorial design with an additional control group was used. The control group judged a trial that consisted of a single burglary charge which we will refer to as the "target" offense. The experimental groups judged a trial that consisted of the same target charge in combination with two other charges that represented the experimental manipulations. The independent variables were (1) charge similarity: identical, similar or dissimilar, (2) evidence similarity: similar or dissimilar and (3) judges' instructions: present or absent. Charge and evidence similarity were crossed factorially and judges' instructions were manipulated for the similar evidence, but not the dissimilar evidence conditions, resulting in 9 experimental groups and one control group. (It was not financially possible to run the complete factorial design using jury pool subjects. Study 3 reports results of the full factorial conducted with undergraduates.)

Based on pretesting of the materials, the independent variables were defined as follows: Charge similarity was defined as the type of crime and the circumstances surrounding the crime, where identical charges were three service station burglaries, all committed in the same manner; similar charges were three somewhat similar burglaries committed at different establishments--the target service station burglary, a house burglary, and burglary of a commercial business establishment, and dissimilar charges were burglary, assault and armed robbery charges.

Evidence similarity was defined as the type of evidence brought to trial by the attorneys to prove their case. For similar evidence conditions the evidence for each charge was circumstantial evidence that the defendant was seen driving suspiciously near the scene around the time of the crime with no explanation for his whereabouts. For dissimilar evidence conditions the main evidence was different for each charge. For example, the same circumstantial evidence for the target offense might be combined with a charge



burglary, assault and armed robbery cases tried in Wisconsin. The target offense was adapted from a complete trial transcript. Pretesting of the materials (Study 1) indicated that they met the requirements of the research in terms of case strength and the independent variables. For each of the thirteen charges, fact sheets containing all the information about the case were prepared. In addition, a fact sheet was prepared for each witness, containing all the information that witness would need to know on the witness stand. Two experienced trial attorneys were recruited to serve as the attorneys in the trial re-enactments. They were given the fact sheets in advance, from which they prepared their own opening statements, closing arguments and questions for witnesses. Witnesses were volunteers recruited from among graduate students, psychology department staff members, and advanced undergraduates. The same person (an advanced undergraduate male) played the part of the defendant for all the trials. Each witness was given the fact sheet beforehand containing all the information he or she would need in testimony. To assure a high level of realism, there were no scripts. Witnesses simply came prepared to answer any questions that might be asked of them by the attorneys. The trial re-enactments were videotaped at the University of Wisconsin Law School courtroom over the course of a weekend. Each of the thirteen offenses was filmed individually (the joinder manipulations were accomplished through editing). The cases were essentially "tried" spontaneously on camera, resulting in an abbreviated but complete trial lasting from 30 to 45 minutes for each individual offense.

The experimental conditions were created by editing together combinations of three charges each, all of which contained the target offense in combination with two other charges. In fact, it is because the content of the target offense remained constant across charges that it was not necessary to precisely control the content of the trial re-enactments. The edited versions were presented in the form a joined trial is actually conducted. Following an introduction by the judge, the prosecuting attorney makes opening statements for each of the three charges, followed by opening statements for each charge by the defense attorney. The prosecuting attorney then calls witnesses for the first, second, and third charge respectively, and each witness is subject to both direct and cross-examination. After the prosecution has called all its witnesses, the defense calls witnesses for each of the three charges. Following testimony, each attorney presents closing arguments for each of the three charges. Finally come the judge's instructions on the law, including the joinder instruction for the three instruction conditions. The content of the target offense remained identical in all conditions, and the target offense material always came first. The target offense presented as a single trial constituted the control group for all experimental conditions. Each joined trial lasted from 1 1/2 to 2 hours, and the single trial lasted approximately 50 minutes.

Procedure. Subjects participated in evening sessions at the psychology building in groups of six to sixteen per session. A block randomization procedure was used to determine the order in which the conditions were run. Subjects first viewed the videotaped trial in black and white on a 19" television monitor. Immediately following the trial they individually answered a short "pre-deliberation" questionnaire on which they indicated their verdict preference, certainty in verdict and likelihood of defendant guilt. While they were completing the questionnaire the experimenter set up a color video camera in the corner of the room for filming deliberations. Subjects

were then formed into one or two groups of six members for deliberation, and any extra subjects were sent to a separate location to begin answering a "non-deliberating" juror questionnaire (described below). When two six-person groups were run in a single session, one group was taken to a second room in which a camera had been previously set up. Juries were instructed to deliberate and reach a unanimous verdict, and were given a form on which to record their verdict for each charge. If an experienced juror had previously been foreman of a jury, that juror was appointed foreman, otherwise a foreman was selected randomly by the experimenter. The experimenter then turned on the camera to begin taping the deliberations, and left the room to allow the group to deliberate in private. A monitor was set up outside the room so that the experimenter could periodically check on the progress of deliberations. A time limit of one hour was placed on deliberations, and juries were not allowed to declare themselves hung if they had not deliberated the full hour. Groups were also given warnings by the experimenter after 50 and 55 minutes of deliberation indicating that they had only a few more minutes in which to reach a verdict. Following deliberations, subjects individually completed a "post-deliberation" questionnaire which is described below. Subjects were then debriefed in a group and paid \$20 by the experimenter. The entire session lasted from 2 to 3 1/2 hours.

#### Dependent Measures

Pre-deliberation questionnaire. Prior to deliberation, jurors individually answered a short questionnaire which contained the following measures for each charge: (1) verdict-- guilty or not guilty, (2) certainty in verdict on a 9-point scale, (3) probability of guilt of the defendant on a 9-point scale and (4) reasonable doubt standard on a 9-point scale. Due to the failure of a large number of subjects to understand the last question, it was excluded from analysis.

Group verdict. After the group had reached a verdict of guilty or not guilty, the foreman recorded the verdict on a sheet provided for each charge. For hung juries an individual poll was taken and recorded on the sheet provided.

Coding scheme for deliberations. Each statement by each juror was coded into at least one of 38 categories. If a statement fit in more than one category it was coded more than once; however, most statements fit only a single category. The categories included:

1. Case facts. Case facts were defined as any statements concerning the facts or evidence in the case, and included descriptions of events that took place, references to facts brought out in testimony, or any factual information about the case. Case facts were also coded for direction, resulting in four categories of case facts: (a) positive (pro-defendant), (b) negative (anti-defendant), (c) neutral and (d) questions about case facts. (For all categories, positive direction was defined as pro-defendant, negative as anti-defendant.)

2. Errors. Factual errors were coded for the target charge only. Errors were defined as statements that were factually incorrect. Four categories of errors were coded: (a) positive, (b) negative, (c) neutral, (d) corrections of errors--these occurred when one group member pointed out



another group member's error.

3. Verdict statements. These were defined as explicit statements of verdict preference in categories (a) guilt, (b) innocence, (c) questions about verdicts (e.g. "do you think he's guilty?"). Included in this category were formal voting or polling of the jurors, as well as any statement in which a juror indicated they thought the defendant was either innocent or guilty. A special category (d) was added under the verdict heading because it came up fairly often, this category applied to statements indicating that a juror thought the defendant was guilty, but that there was not enough evidence to prove it, i.e., "I think he's guilty but the evidence doesn't prove it," and variations thereof.

4. Reasonable doubt. This heading defined statements referring to the legal definitions of reasonable doubt and presumption of innocence as presented in the judges instructions at the end of the trial. These statements were coded into three categories: (a) positive--indicating that there was a reasonable doubt, (b) negative--indicating lack of doubt, and (c) neutral--for example, discussion of what the judge meant by reasonable doubt.

5. Verdict elements. Verdict elements are defined as statements referring to the part of the judges instructions in which the charge is defined. For example, in defining burglary the judge instructs the jury that in order to prove burglary it must be determined that the defendant entered the building, without the owner's consent and with intent to steal. Verdict elements were coded as (a) positive--indications that various elements of the charge had been proved (b) negative--indications that certain elements had not been proved, or (c) neutral--which involved discussion of what the definition for a particular charge was, with no positive or negative connotations.

6. Evidence sufficiency. These were statements indicating that there was (negative) or was not (positive) enough evidence to convict the defendant.

7. Witnesses. Evaluative statements about trial witnesses were coded into six categories: prosecution witness (positive or negative), the defendant (positive or negative) and other defense witnesses (positive or negative). There were almost no statements in any of these categories--when the number of statement categories, all six categories were summed, they constituted less than 1% of all statements. Therefore these data were not analyzed and will not be reported.

8. Joinder issues. Statement specifically referring to joinder were coded for joined conditions only, since subjects in the control condition would have no reason to make such references. Five categories of joinder issues were coded. (a) Confusion was defined as statements which indicated that jurors did not remember which facts went with which charge, as well as direct statements that it was confusing to have the charges joined together. (b) Accumulation was defined as using evidence from one charge in judgements of other charges. For example, jurors might mention facts from one offense when discussing another, or draw connections among the charges. (c) Criminal inferences were statements that the defendant must be guilty, since he was charged with multiple offenses. (d) Instructions were coded for the instructions condition only--these were references to the joinder instructions manipulation. There were so few statements in this category (0.1%) that these

were not analyzed. (e) Other joinder issues were defined as references to joinder that were not codable in any of the above categories.

9. Other statements. Other statements coded were (a) references to the experiment, (b) directions designed to structure the discussion, (c) group outbursts, (d) irrelevant statements, and (e) uncodable statements. Category (d) and (e) were later collapsed into a "miscellaneous" category.

Post-deliberation questionnaire. Following deliberations, jurors responded to a longer questionnaire designed to assess the processing of trial information. Non-deliberating subjects completed the same questionnaire without participating in deliberations.

1. Background information. Subjects were asked to indicate their age, sex, occupation, income, and various other demographic characteristics.

2. Trial ratings. Subjects rated the realism of the trial and their interest and involvement in the trial on 9-point scales.

3. Requirements of proof. Subjects were asked two questions designed to assess their perceptions of innocence and reasonable doubt. The first question was: "A defendant should be found guilty if there is at least \_\_\_% chance he has committed the crime." This can be considered a measure of a juror's criterion for conviction. A second question was: "Approximately \_\_\_% of all defendants brought to trial are guilty." This can be considered a rough measure of the juror's presumption of innocence.

4. Defendant ratings. Subjects were asked to rate the defendant on the following eleven 9-point bipolar scales: honest-dishonest, dangerous-not dangerous, likeable-dislikeable, good-bad, sincere-insincere, believable-unbelievable, calm-nervous, moral-immoral, attractive-unattractive, future crime likely-unlikely, a typical criminal-not a typical criminal. The purpose of these measures was to assess inferences about the defendant's disposition.

5. Memory. Subjects were given two memory tasks designed to assess their degree of confusion between charges. (a) Free recall. For each case subjects were asked to list the evidence that most strongly supported the prosecution's case, and to do the same for defense evidence. (b) Recognition. Subjects were given a multiple choice recognition task and were asked to choose which facts were contained in the target charge from among (1) correct items actually contained in the target offense, (2) factual errors about the target offense and (3) items from the other two non-target charges which were attributed to target offense witnesses. These items constituted a measure of confusion. (c) Judges' instructions. Subjects in the instructions conditions were also asked for free recall of the judge's instructions with respect to multiple charges.

6. Evidence ratings. Subjects were asked to rate the strength of the evidence for prosecution and defense overall for each charge on a scale from 1 (very weak) to 9 (very strong). Subjects were also asked to rate the incriminating value of four specific items of evidence for each charge, two for prosecution and two for defense, on a scale from 1 (strongly indicates innocence) to 9 (strongly indicates guilt). The evidence ratings provided a

measure of accumulation of evidence.

7. Similarity ratings. As a manipulation check, subjects in joined conditions rated the similarity of the charges and the evidence contained in the three charges. Subjects also rated the similarity of the charges with respect to the legal precedents for joining charges. They were asked to what extent the three charges established a similar motive, intent, plan, identity, and disposition. On the final page of the questionnaire subjects were asked questions specifically pertaining to joinder in terms of fairness and whether they felt that joinder had affected their decisions.

#### Pre-deliberation Results

Manipulation Checks. Subjects in joined conditions rated the similarity of the type of offense charged and the evidence contained in the charged offenses on scales from 1 (highly similar) to 9 (highly dissimilar). Data for each measure were analyzed in 3 (charge similarity) x 2 (evidence similarity) analyses of variance collapsed across the instruction variable. For charge similarity ratings there was a main effect for charge similarity,  $F(2,617) = 48.75$ ,  $p < .0001$ , and a main effect for evidence similarity,  $F(1,617) = 6.83$ ,  $p < .01$ , and no interactions. The mean ratings of charge similarity in identical, similar, and dissimilar charge conditions were 3.39, 4.17, and 5.62 respectively, and this effect was fairly large in magnitude, Effect Size (Epsilon, Cohen, 1977) = .37. Subjects also rated charges containing similar evidence as more similar ( $M = 4.22$ ) than charges containing dissimilar evidence ( $M = 4.76$ ), although this effect was not large,  $ES = .10$ . For evidence similarity ratings there was a main effect for evidence similarity,  $F(1,617) = 20.53$ ,  $p < .0001$ ,  $ES = .175$ , no effect for charge similarity, and no interactions. Subjects in similar evidence conditions rated the evidence as more similar ( $M = 3.95$ ) than subjects in dissimilar evidence conditions ( $M = 4.75$ ). The results indicate that both manipulations were effective, although the charge similarity manipulation was stronger than the evidence similarity manipulation.

Verdicts. Prior to deliberation, subjects provided individual verdict preference (guilty or not guilty), and rated certainty in verdict and probability of the defendant's guilt on 9-point scales. These ratings were analyzed for all subjects (deliberating and non-deliberating), since deliberating and non-deliberating groups were equivalent prior to deliberation. Analyses were performed on the first (target) charge only, since it was the only charge that remained constant across conditions--the other two charges served as the experimental manipulations. Because the design was not a full factorial design, and since many of the effects of interest involved comparison of different experimental conditions with the single control group, a series of planned comparisons was performed as recommended by Himmelfarb (1975). A modified Bonferroni procedure was used to control error rates (Keppel, 1982). Results with a probability less than .035 are considered significant. Experimental-control comparisons for which the hypotheses predicted higher ratings in joined versus control conditions were tested using one-tailed significance tests, and experimental-control comparisons for which the hypotheses were not directional used two-tailed tests. We will confine our discussion here to the effects of primary interest: joinder (versus the control group), instructions (versus no-instructions), charge similarity, evidence similarity, and interactions

among these factors.<sup>1</sup>

The proportion of individual guilty verdicts and probability of guilt ratings obtained in the ten experimental conditions are provided in Table 2-4. Analysis revealed a significant effect for joinder,  $t(722) = 2.57$ ,  $p < .01$ , Effect Size = .10, with a mean proportion of 39% guilty verdicts in joined conditions, as opposed to 24% guilty verdicts in the control group. There was no effect for instructions; in fact, conviction rates in joined-instructions ( $M = .38$ ) and no-instructions ( $M = .39$ ) conditions were virtually identical. There were no effect for charge similarity or evidence similarity, although there was a tendency of marginal significance,  $F(1,722) = 3.05$ ,  $p = .08$ ,  $ES = .06$  for more convictions in dissimilar evidence ( $M = .43$ ) than similar evidence ( $M = .35$ ) conditions. There were no interactions among any of the variables. On the basis of the verdict results, it can be concluded that joinder significantly increased the likelihood of conviction, and that judges' instructions were totally ineffective in reducing these effects.

Certainty in verdict. In analyses of verdict certainty ratings, no significant results were obtained. All subjects were equally confident in their verdicts regardless of condition. Mean certainty ratings ranged from 6.69 to 7.47 on a 9-point scale.

Probability of guilt. It was predicted that subjects in joined conditions would judge the defendant as more likely to be guilty than subjects in the single case control group. Although the direction of the means supports this prediction, the joinder effect overall was only marginally significant,  $t(721) = 1.46$ ,  $p = .07$ ,  $ES = .05$ , with means of 5.41 for joined trials and 4.99 for the control group. There was a marginally significant effect for evidence similarity,  $F(1,721) = 3.85$ ,  $p = .05$ , in which subjects in dissimilar evidence conditions rated guilt as higher ( $M = 5.69$ ) than subjects in similar evidence conditions ( $M = 5.21$ ). This result parallels the result obtained for verdict--subjects also returned more guilty verdicts when evidence was dissimilar. There were no instruction effects or interactions on the probability of guilt ratings.

#### Deliberation Results

Reliability checks. The deliberation videotapes were coded by two undergraduate assistants. One person coded approximately two-thirds of the juries, the other coded the remaining one-third. In order to assess coder reliability, the two coders independently coded a sample of four juries (approximately 2 hours of deliberation). Correlations between coders were computed on the number of statements coded for each juror under various category headings. The correlations were as follows: case facts:  $r = .85$ ; verdict statements:  $r = .86$ ; other legal issues (reasonable doubt, verdict elements, evidence sufficiency):  $r = .79$ ; errors:  $r = .52$ ; multiple charges:  $r = .57$ ; other statements (directions, experiment, miscellaneous):  $r = .90$ . Thus, with the exception of the error and multiple charge categories, (both of which were infrequent) interrater reliability was high. Also, close to 90% of all statements were coded into meaningful categories, with only 11.5% coded into the miscellaneous category.

We began our deliberation analyses by examining discussion of the target offense (which was the only offense common to all joined conditions). For each juror, the frequency of statements in each category was divided by the total number of statements made by that juror. Analyses were performed on the proportion of statements made in each category. The mean proportion of statements on each category are presented for each experimental condition in Table 2-6 and for each charge (first, second and third) in Table 2-7. These are broken down by experimental conditions for the most important categories. Planned comparisons were used to examine the following effects: Joinder effects (control vs. joined conditions), instruction effects (vs. no instructions), charge similarity effects, evidence similarity effects and interactions.

Case facts. There were two marginally significant joinder effects for case fact categories, one for negative facts,  $t(530) = 1.75$ ,  $p = .08$ ; and one for questions about case facts,  $t = 1.86$ ,  $p = .06$ . Negative facts were more frequent in joined conditions ( $M = 10.9$ ) than in the control group ( $M$

<sup>1</sup> A complete summary of all analyses is available from the authors.

Table 2-4. Proportion of guilty verdicts

CONTROL GROUP	EVIDENCE SIMILARITY	CHARGE SIMILARITY			
		IDENTICAL	SIMILAR	DISSIMILAR	
.24 n = (83)	Similar	.32 (77)	.36 (68)	.33 (74)	No Instruc- tions
	Dissimilar	.43 (68)	.41 (72)	.46 (72)	
	Similar	.36 (69)	.43 (76)	.35 (69)	Instruc- tions

Table 2-5. Proportion of statements in each category (target charge).

	Grand Mean	Condition (see Table 2-3)										Joined Mean
		Control 0	1	2	3	4	5	6	7	8	9	
n=	(540)	(42)	(60)	(60)	(48)	(54)	(60)	(48)	(60)	(54)	(54)	(498)
<u>Case Facts</u>												
Positive	11.7	14.5	13.1	9.4	7.8	10.0	11.9	14.4	10.1	12.7	11.9	11.5
Negative	10.6	6.8	10.4	11.8	9.3	9.0	13.2	11.1	7.0	17.3	12.8	10.9
Neutral	7.8	7.1	9.2	9.3	4.4	7.4	6.1	5.2	8.3	10.5	7.7	7.8
Questions	9.7	13.0	8.1	9.8	8.4	8.5	10.4	8.7	7.8	11.6	10.8	9.4
<u>Errors</u>												
Positive	0.2	0.5	0.2	0.2	0.0	0.3	0.2	0.0	0.1	0.0	0.5	0.2
Negative	0.2	0.3	0.1	0.3	0.2	0.5	0.4	0.2	0.0	0.3	0.0	0.2
Neutral	0.6	0.8	0.8	1.0	0.8	0.1	0.6	0.2	0.8	0.8	0.4	0.6
Corrections	0.8	0.4	0.7	1.5	0.7	0.4	1.2	0.4	1.2	1.1	0.7	0.9
<u>Verdict Statements</u>												
Guilty	6.4	2.7	6.0	5.8	3.9	6.0	3.4	2.0	1.3	3.3	6.9	7.7
Not guilty	14.4	11.8	9.9	13.4	29.5	14.4	13.3	10.4	17.5	11.6	14.1	14.6
Questions	1.2	0.7	0.6	0.7	0.9	1.3	1.3	2.8	1.3	2.1	0.5	1.2
Guilty but not enough proof	1.5	2.5	0.5	2.0	1.9	0.9	0.9	1.5	1.9	1.0	1.7	1.4
<u>Reasonable Doubt</u>												
Positive	1.9	3.8	1.6	1.1	0.8	0.9	2.6	1.5	3.5	2.0	1.4	1.8
Negative	0.5	0.2	0.7	0.2	0.2	0.7	0.3	1.3	0.5	0.6	0.3	0.5
Neutral	1.1	2.2	1.3	1.1	0.4	0.7	1.9	0.6	1.1	0.7	0.8	1.0
<u>Evidence sufficiency</u>												
Positive	2.6	3.9	3.5	2.9	1.3	1.5	2.5	3.7	3.6	1.6	1.2	2.5
Negative	0.9	0.6	0.8	1.4	1.1	1.2	1.7	0.7	0.3	1.1	0.6	0.9
<u>Verdict elements</u>												
Positive	1.5	1.2	0.8	1.4	0.2	1.2	2.6	3.1	2.1	0.6	1.8	1.6
Negative	0.04	0.05	0.2	0.03	0.0	0.0	0.05	0.0	0.06	0.04	0.0	0.04
Neutral	0.6	0.3	0.6	0.7	0.0	1.1	0.6	1.2	0.9	0.2	0.6	0.7
<u>Joinder issues</u>												
Confusion	0.8	0.0	1.3	1.1	1.0	0.6	0.7	0.6	1.7	0.4	0.2	0.9
Accumulation	0.8	0.0	2.0	1.7	0.05	1.4	0.5	0.2	1.1	0.4	0.0	0.9
Crim. inferences	0.9	0.0	1.1	1.5	0.9	1.5	0.3	0.3	0.9	0.1	0.0	1.0
Other	0.9	0.0	2.5	1.2	0.6	0.6	0.6	0.1	1.4	0.5	0.2	1.0
<u>Other (Overall means only)</u>												
Witnesses	0.8											
Experiment	4.3											
Directions	6.2											
Group outburst	0.9											
Miscellaneous	10.6											

Table 2-6

## Proportions and frequencies of statements per juror in each category.

Category	Charge 1		Charge 2		Charge 3		Total	
	%	Freq	%	Freq	%	Freq	%	Freq
<b>CASE FACTS</b>								
Positive	11.7	1.82	10.9	0.73	8.9	0.48	11.7	3.22
Negative	10.7	1.66	4.7	0.31	2.5	0.13	8.4	2.31
Neutral	7.6	1.18	8.4	0.56	8.3	0.45	8.4	2.31
Questions	9.6	1.49	8.6	0.57	9.5	0.51	10.4	2.83
Total	39.6	6.15	32.6	2.18	29.2	1.57	38.9	10.70
<b>ERRORS<sup>1</sup></b>								
Positive	0.2	0.03						
Negative	0.2	0.03						
Neutral	0.6	0.09						
Corrections	0.9	0.14						
Total	1.9	.29						
<b>VERDICTS</b>								
Guilty	6.5	1.01	0.4	0.03	2.5	0.13	4.6	1.27
Not Guilty	14.3	2.22	33.4	2.23	32.7	1.76	18.6	5.12
Questions	1.2	0.19	0.7	0.05	0.6	0.03	0.9	0.25
Guilty, No Proof	1.5	0.23	1.0	0.07	0.7	0.04	1.2	0.33
Total	23.5	3.65	35.5	2.37	36.5	1.97	25.3	6.96
<b>REASONABLE DOUBT</b>								
Positive	2.0	0.31	0.9	0.06	0.8	0.04	1.6	0.44
Negative	0.5	0.08	0.1	0.01	0.1	0.01	0.3	0.08
Neutral	1.1	0.17	0.3	0.02	0.6	0.03	0.9	0.25
Total	3.6	0.56	1.3	0.09	1.5	0.08	0.8	0.77
<b>EVIDENCE SUFFICIENCY</b>								
Positive	2.6	0.40	3.1	0.21	2.5	0.13	2.7	0.74
Negative	0.9	0.14	0.1	0.01	0.0	0.00	0.6	0.17
Total	3.5	.54	3.2	0.22	2.5	0.13	3.3	0.91
<b>VERDICT ELEMENTS</b>								
Positive	1.6	0.25	1.3	0.09	0.7	0.04	1.4	0.39
Negative	0.0	0.00	0.2	0.01	0.0	0.00	0.1	0.03
Neutral	0.7	0.11	0.7	0.05	0.6	0.03	0.7	0.19
Total	2.3	0.36	2.2	0.15	1.3	0.07	2.2	0.61
<b>MULTIPLE CHARGES</b>								
Confusion	0.8	0.12	1.0	0.07	0.9	0.05	1.0	0.28
Accumulation	0.8	0.12	0.9	0.06	0.5	0.03	1.0	0.28
Inferences	0.9	0.14	1.0	0.07	0.8	0.04	1.0	0.28
Other	0.9	0.14	1.0	0.07	2.3	0.12	1.4	0.39
Total	3.4	0.53	3.9	0.26	4.5	0.24	4.4	1.21
<b>OTHER</b>								
Experiment	4.3	0.67	4.1	0.27	6.5	0.38	4.6	1.27
Directions	6.4	0.99	3.8	0.25	4.4	0.24	5.0	1.38
Outburst	1.0	0.16	0.8	0.05	1.0	0.05	1.0	0.28
Irrelevant, Uncodable	10.6	1.65	9.1	0.61	12.5	0.68	11.5	3.16
Total	22.3	3.46	17.8	1.19	24.4	1.31	22.1	6.08
SUM	100	15.52	100	6.69	100	5.38	100	27.51

<sup>1</sup>Errors were not coded for Charges 2 and 3.

= 6.8), and there were more questions in the control group (M = 13.0) than in joined conditions (M = 9.4). Although the joinder effect overall for positive facts was not significant, there were significantly fewer facts in the joined-similar charge conditions (M = 8.9) than in the control group (M = 14.5),  $t = 2.08$ ,  $p = .04$ . Thus there is some indication that subjects made more negative and fewer positive statements when charges were joined. For neutral case facts, there were no joinder effects, but there was an instruction effect,  $t = 1.92$ ,  $p = .05$ , which indicated that subjects made more neutral statements with instructions (M = 8.8) than without (M = 6.9). There was also a main effect for charge similarity on neutral case facts,  $F(2, 306) = 3.30$ ,  $p = .04$ , which indicated that jurors made more neutral statements in identical (M = 9.2) than in similar (M = 6.3) or dissimilar (M = 6.1) charge conditions. Evidence similarity had no effects on any case facts categories, and neither charge similarity nor instructions influenced positive facts, negative facts or questions. There were no interactions among any of the variables.

**Errors.** None of the analyses yielded significant effects on any of the error categories, and jurors made few errors at all--the four error categories together constituted only 1.8% of target charge discussion.

**Verdict statements.** Overall there were more not guilty (M = 6.4) statements--this parallels the initial and final verdict results in which more not guilty than were obtained. For the proportion of guilty statements, the joinder effect overall was not significant, but there was a marginally significant difference between the joinder-identical charge conditions and the control group,  $t = 1.75$ ,  $p = .08$ , with more guilty statements in the joined-identical conditions (M = 5.9) than the control group (M = 2.7). In terms of not guilty statements, the joinder effect was again not significant, but there was a significant difference between joined-similar charge conditions and the control group,  $t = 2.71$ ,  $p = .007$ , which was opposite to predictions--subjects in the joined-similar charge conditions made more not guilty statements (M = 22.0) than subjects in the control group (M = 11.8). There was also a main effect for charge similarity,  $F(2, 306) = 10.04$ ,  $p = .0001$ , not guilty statements in the similar charge condition were also more frequent than in identical (M = 10.7) or dissimilar (M = 11.8) charge conditions. There were no significant effects for any of the manipulations on verdict questions or the "guilty but no proof" category. There were no evidence similarity effects, instruction effects or interactions on any of the verdict categories.

**Reasonable doubt.** For positive reasonable doubt statements (i.e., expressions of doubt) there was a significant joinder effect,  $t = 2.59$ , with more doubt expressed in the control (M = 3.8) versus Joined (M = 1.8) conditions. There was also a main effect for instructions on positive reasonable doubt,  $t = 1.97$ ,  $p = .05$ ; subjects expressed more doubt with instructions (M = 2.3) than without (M = 1.4). There was also a charge similarity effect on positive reasonable doubt,  $F = 2.97$ ,  $p = .05$ , with means of 1.3, 0.7 and 2.3 in identical, similar, and dissimilar charge conditions. Thus subjects expressed most doubt when charges were dissimilar. There were no evidence similarity effects or interactions for positive reasonable doubt. The negative reasonable doubt (i.e., lack of doubt) category occurred infrequently (0.5%) and none of the analyses yielded significant effects. For neutral reasonable doubt (i.e., discussions of the definition of reasonable doubt), there was a significant joinder effect,  $t =$

2.37,  $p = .02$ ; neutral statements occurred more frequently in the control group ( $M = 2.2$ ) than in experimental groups ( $M = 1.0$ ). There were no other main effects or interactions for neutral reasonable doubt statements.

Evidence sufficiency. The overall joinder effect on positive evidence sufficiency (i.e., lack of evidence) was not significant; however, the similar charge-joined conditions were significantly different from the control group,  $t = 2.24$ ,  $p = .025$ . Subjects in the control group made more statements concerning the lack of sufficient evidence ( $M = 3.9$ ) than subjects in the joined-similar charge conditions ( $M = 1.4$ ). There were no other main effects or interactions in the evidence sufficiency category (positive or negative).

Verdict elements. For positive verdict elements, the joinder effect overall was not significant. There was a significant difference between the joined-dissimilar charge condition ( $M = 2.8$ ) and the control group ( $M = 1.2$ ),  $t = 2.14$ , as well as a significant charge similarity effect,  $F = 7.76$ ,  $p = .0005$ , with more positive statements in dissimilar ( $M = 2.8$ ) than similar ( $M = 0.8$ ) or identical ( $M = 1.2$ ) charge conditions. No other analyses yielded significant effects or positive verdict elements, and there were no significant effects on negative verdict elements (which were extremely rare  $M = 0.047$ ). For neutral verdict elements, there was a main effect for evidence similarity,  $F = 4.58$ ,  $p = .03$ , with means of 1.1 for dissimilar evidence and 0.5 for similar evidence conditions. No other effects or interactions were significant.

Joinder issues. Because discussion of joinder issues was only possible in joined conditions, it was not meaningful to compare joined and control groups. For statements indicating confusion, there were no effects for charge similarity, evidence similarity or instructions. For accumulation (combining evidence across charges) there was a main effect for charge similarity,  $F = 4.76$ ,  $p = .01$ , with means of 0.3, 0.7, and 1.8 in dissimilar, similar and identical charge conditions. This suggests that accumulation increased as charge similarity increased, although this result should be interpreted with caution due to the lower reliability of ratings in the joinder issue categories. For criminal inferences, there was a main effect for evidence similarity,  $F = 4.91$ ,  $p = .03$ , with more inferences in dissimilar ( $M = 1.8$ ) than similar ( $M = .07$ ) evidence conditions. For other joinder statements (that did not fit into confusion, accumulation, or criminal inference categories) there was a main effect for charge similarity,  $F = 4.69$ , with means of 0.9, 0.6, and 1.8 in dissimilar, similar and identical charge conditions. No other effects were significant on joinder issues.

Taken as a whole, the effects of manipulations on the proportion of statements in various categories were relatively weak. This was particularly true for the charge similarity, evidence similarity, and instruction manipulation, which produced almost no influence. However, there were some effects obtained for joinder, and these were consistent with predictions and with questionnaire results. Joinder effects occurred primarily in two categories: case facts and reasonable doubt. Subjects in joined conditions made less positive and more negative statements than subjects in the control group, and expressed less doubt about the defendant's guilt.

Individual and group voting behavior. Table 2-7 presents the group verdicts for each of the three charges as a function of the number of jurors who initially voted to convict. It is apparent that all cases were on the weak side, with 63% acquittals on the first charge, 80% acquittals on the second, and 79% on the third. The results indicate that majorities tended to prevail; on the first charge there were only six reversals of initial majorities; two groups with initial 4-2 splits for conviction ultimately acquitted, and four groups with 2-4 splits convicted. For the second and third charges, there were four reversals apiece. Thus, the present data demonstrate the well documented finding that the initial juror vote distribution is a good predictor of the final group outcome (Davis, 1973; Kalven & Zeisel, 1966; Penrod & Hastie, 1979, 1980; Stasser & Davis, 1981).

Deliberations across time. In order to investigate the progress of deliberations over time, the frequencies of statements in each category were analyzed in separate 3 X 4 (Charge X Time) repeated measures analyses of variance, where charge refers to discussion of the first, second, and third charges respectively; and time refers to the first, second, third, and fourth quarters of deliberation. A similar, and not very surprising, pattern of results was obtained on most categories, so we will concentrate out discussion here on the most important categories only (case facts and verdicts).

For positive case facts, there was a main effect for Charge,  $F(2,870) = 60.37$ ,  $p < .001$ ; a main effect for Time,  $F(3,1305) = 3.13$ ,  $p = .023$ ; and a Charge X Time interaction,  $F(6,2610) = 17.62$ ,  $p < .001$ . The charge effect indicated that positive facts were most frequently mentioned with respect to the first charge ( $M=1.97$ ), and less frequently mentioned for the second ( $M=.89$ ) and third ( $M=.63$ ) charges. Similar charge effects were obtained on virtually all categories, reflecting the fact that more time was spent discussing the first charge than the second two. For the time main effect, the frequencies during the 1st, 2nd, 3rd, and 4th quarters of deliberation were 0.86, 0.80, 1.0, and 0.83. The Charge X Time interaction is graphed on the left side of Figure 2-1. Not surprisingly, discussion of Charge 1 was most frequent at Time 1 and declined steadily over time. Discussion of positive facts for Charge 2 first increased and then decreased, while Charge 3 was primarily discussed during the second half of deliberations. This suggests that jurors discussed the three cases in logical progression, beginning with the first charge, then moving on to the second and finally the third. A similar pattern of results was obtained on most content categories. This finding supports Hastie, Penrod, and Pennington's (1983) notion of agenda setting, which postulates that juries form agendas, or decision sequences, which they follow when making multiple judgments.

For negative case facts, there were again main effects for Charge,  $F(2,870) = 93.78$ ,  $p < .001$ ; Time,  $F(3,1305) = 12.29$ ,  $p < .001$ ; and a Charge X Time interaction,  $F(6,2610) = 27.63$ ,  $p < .001$ . The frequencies for Charges 1, 2, and 3 respectively were 1.86, 0.48, and 0.20; and the frequencies in the 1st, 2nd, 3rd, and 4th quarters were 0.84, 0.74, 1.34,

TABLE 2-7

Jury Verdicts as a Function of Initial Votes

Initial votes for conviction	CHARGE 1 VERDICTS			CHARGE 2 VERDICTS			CHARGE 3 VERDICTS		
	Not Guilty	Guilty	Hung	Not Guilty	Guilty	Hung	Not Guilty	Guilty	Hung
0	8	0	0	15	0	0	30	0	2
1	19	0	3	31	0	1	28	1	1
2	22	4	4	26	1	2	13	1	0
3	12	2	7	5	1	2	6	2	1
4	2	8	4	2	1	2	2	1	1
5	0	4	0	1	0	0	0	1	0
6	0	1	0	0	0	0	0	0	0
Sum	63	19	18	80	3	7	79	6	5

and 0.47. The Charge X Time interaction is graphed in the right half of Figure 2-1, and the pattern of results is very similar to that obtained for positive facts. Both charge and time main effects as well as their interaction were statistically significant for neutral and question case fact categories, and the pattern of results was again similar (these results are not reported here, but can be obtained from the authors).

Repeated measures analyses were also performed on jurors' verdict preferences. For guilty votes, there were main effects for Charges,  $F(2,874) = 83.98, p < .001$ ; Time,  $F(3,1311) = 11.55, p < .001$ ; and a Charge X Time interaction,  $F(6,2622) = 15.03, p < .001$ . The mean frequencies of guilty statements per juror were 0.74 for Charge 1, 0.28 for Charge 2, and 0.17 for Charge 3. The frequencies in the 1st, 2nd, 3rd and 4th quarter were 0.40, 0.22, 0.23, and 0.34. The Charge Time X interaction is graphed in Figure 2-2, left side. For the first Charge, guilty statements were most frequent in the first quarter and then leveled off, whereas for the second and third charges, guilty statements increased across time. For not guilty preferences, similar results were obtained. There were main effects for Charges,  $F(2,874) = 5.25, p = .005$ ; Time,  $F(3,1311) = 37.43, p < .001$ ; and a Charge X Time interaction,  $F(6,2622) = 28.28, p < .001$ . The mean frequencies for Charges 1, 2, and 3 were 1.13, 1.17 and 1.02; and the frequencies in the first, second, third, and fourth quarters respectively were 0.72, 0.67, 0.66, and 1.28. The Charge X Time interaction is graphed in the right half of Figure 2-2. For Charge 1, not guilty votes decreased from Time 1 to Time 3, but increased somewhat during the last quarter of deliberations. For Charges 2 and 3, not guilty statements increased steadily as time progressed.

Although the verdict and case fact results were similar in most respects, there were some interesting differences between these two categories with respect to the effects of time. Discussion of case facts was concentrated more in the middle of deliberations (reaching its peak in the third quarter), whereas verdict preferences were stated more in the beginning, and even more so at the end. This pattern of results suggests a decision making sequence or agenda (Hastie et al., 1983) in which jurors initially state their verdict preferences, then get down to discussing the facts, and finally review their revised verdict choices and make their decisions.

Total deliberations. Path analyses on each individual charge suggest that for the first charge, jurors' deliberations stem from their initial verdict preferences, and thereby influence the outcome. For the second and third charges, jurors' final votes are based directly on their initial vote preferences, and less on the evidence. We also found that jurors spent less time discussing the second and third charges than they did the first. In addition, the predictive accuracy of the path models for Charges 2 and 3 was far inferior to that of Charge 1. All of the above facts together suggest that jurors may use other information in their judgements of later charges that is not captured in the individual charge models. Specifically, jurors may use information from the charges they judged previously.

Figure 2-1. CASE FACTS

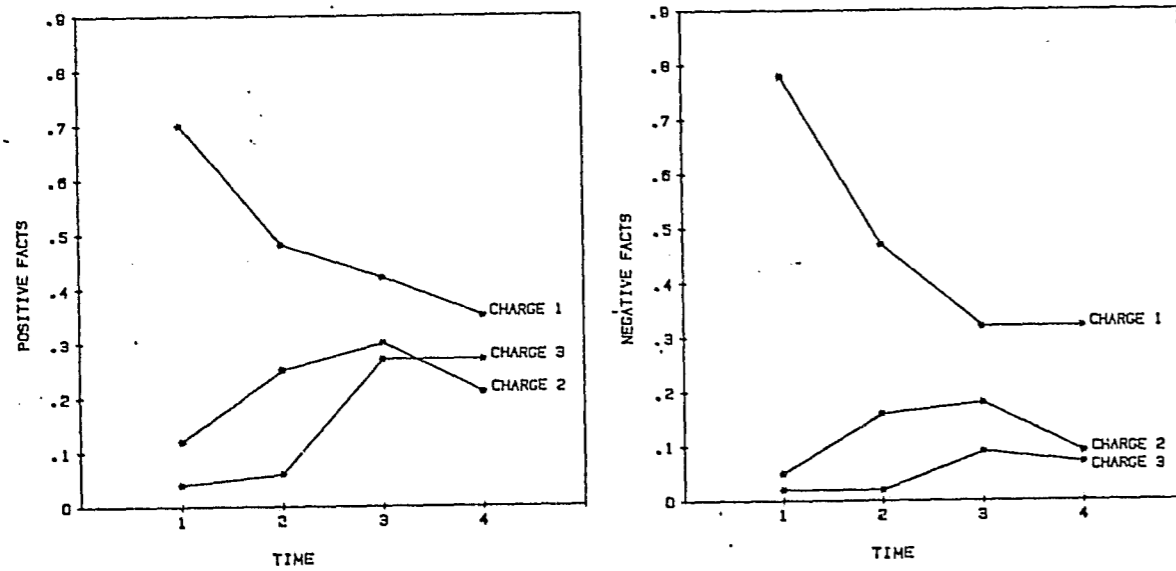
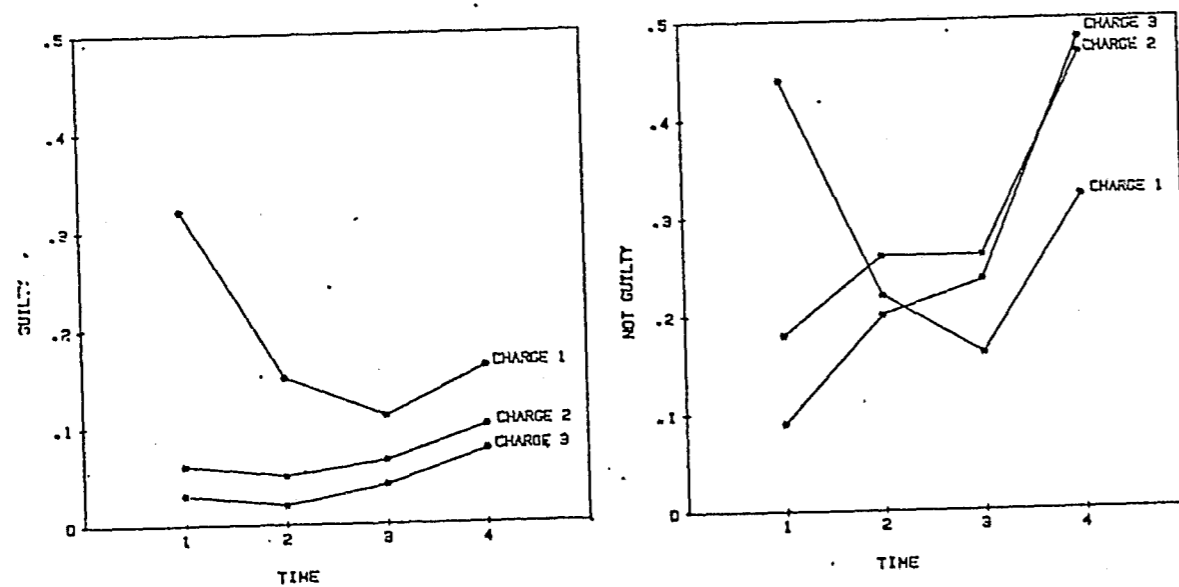


Figure 2-2. VERDICT STATEMENTS



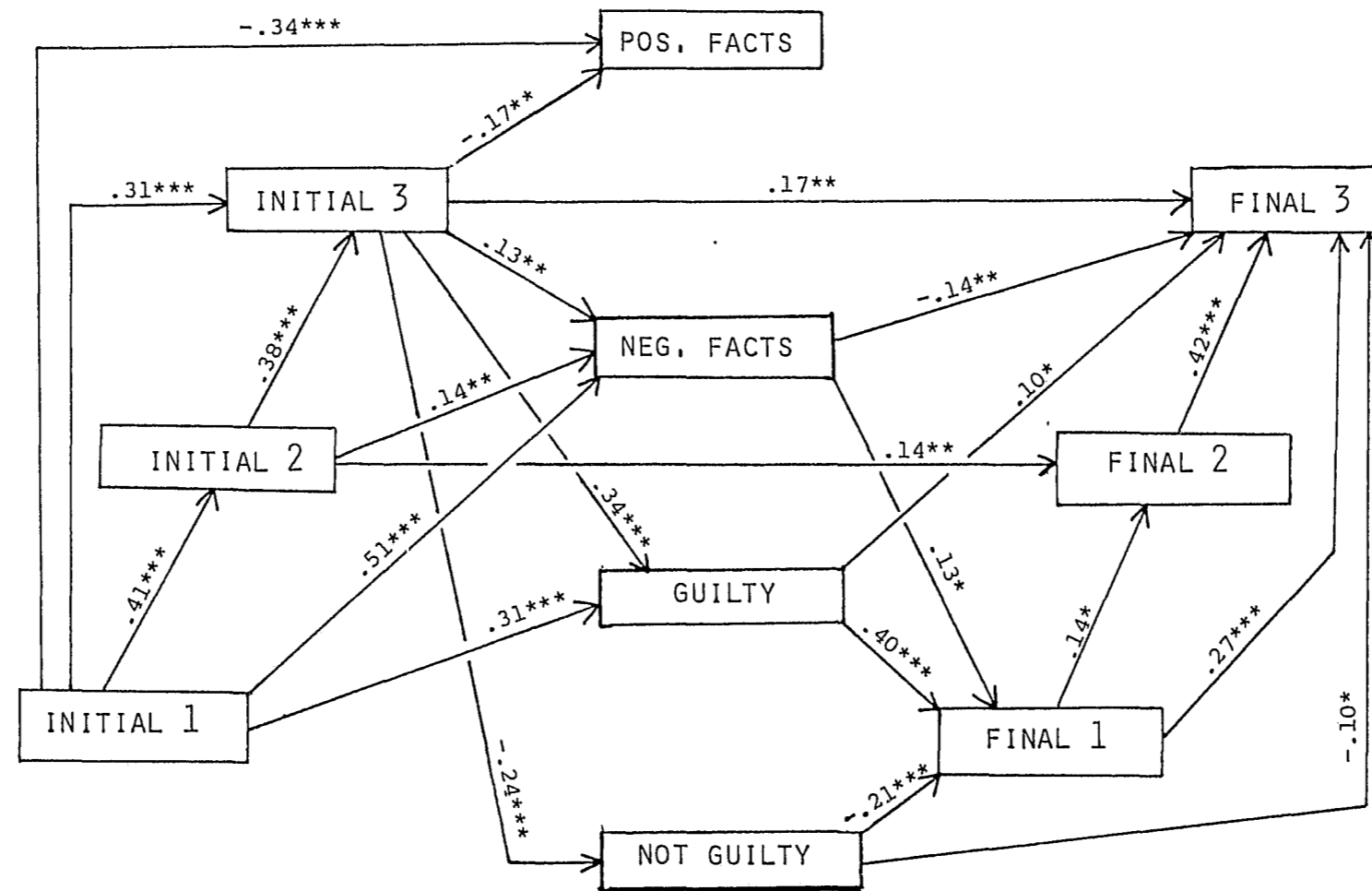
In order to investigate this possibility, we incorporated data from each of the three charges into a single path analysis. The path model included the following variables: jurors' initial votes on charges 1, 2, and 3; the total proportion of statements (collapsed across charges) in each of the four main content categories (positive and negative facts, guilty and not guilty statements), and jurors' final votes on Charges 1, 2, and 3. It was assumed that each decision could influence the subsequent one, therefore the variables were entered into the equation in the following order: initial vote for Charge 1, initial vote for Charge 2, initial vote for Charge 3, deliberation categories (all entered at the same level) final vote for Charge 1, final vote for Charge 2, final vote for Charge 3.

The path analysis results are presented in Figure 2-3. In terms of initial verdict preferences, jurors' votes on all three charges were significantly related to each other. With respect to predicting deliberation content, the initial vote on Charge 1 was a significant predictor of all four categories, the initial vote on Charge 3 significantly predicted all categories except not guilty statements, and the only significant effect for Charge 2's initial vote was on negative facts. Thus, as in the analyses of individual charges, vote 1 had a stronger influence on deliberations than votes 2 and 3.

The final vote for Charge 1 was significantly predicted by 3 of 4 content categories (surprisingly, positive facts had no effect). None of jurors' initial votes significantly influenced final votes for Charge 1; as in the individual case analysis the effects of initial votes were indirect rather than direct. For Charge 2's final vote, none of the content categories were significant predictors, although the coefficients for guilty and not guilty statements were marginally significant ( $p < .10$ ). The only significant predictors of Charge 2 final votes were jurors' initial votes on Charge 2, and their final votes on Charge 1. For Charge 3, 3 of 4 content categories significantly predicted final votes, although these effects were weaker than those obtained for Charge 1. Jurors' final votes on the previous two charges were much stronger predictors of final votes on Charge 3 than was deliberation content, and the initial vote on Charge 3 was also a significant predictor of the final vote. The amount of variance accounted for in the final vote (39%) was far superior to the R-squared of .16 that was obtained when Charge 3 was considered alone.

The preceding analysis creates an overall picture of the dynamics of the deliberation process. Jurors spend a good deal of time discussing the first charge, their deliberations are influenced by their initial votes, and they in fact make their decisions on the basis of their deliberations. Thus, it appears that an informal influence process operates to affect decisions on the first charge. For the second and third charges, jurors spend much less time discussing the case, and do not base their decisions as much on deliberations as they do on their previous votes (and presumably, the votes of other jurors). Thus, it appears that a normative social influence process is in operation for decisions on the second and third charges (Deutsch & Gerard, 1955).

Figure 2-3. Percentage of Total Speaking Time for 4 main Categories/Initial and Final Votes For Each Charge



$F(9,420) = 30.06, p < .001, R^2 = .39$



Although our analyses suggest that jurors tend to base their subsequent decisions on previous ones, there is a plausible alternative explanation for the results, which can be interpreted using the same theory of social influences. The group verdict results presented in Table 2-5b indicate that jurors were more strongly pre-disposed towards innocence for Charges 2 and 3 than they were for Charge 1. Thus, pressure to conform to the majority (i.e., normative influence) was probably stronger for Charges 2 and 3, therefore there was less consideration of the evidence.

#### The Influence of Deliberations on Cognitive Processes

In analyses reported earlier, we found that deliberating and non-deliberating jurors did not differ systematically in their memory for evidence, ratings of the evidence, or ratings of the defendant. A related question concerns the influence of persuasion that takes place during deliberations on jurors' thought processes following deliberations. If jurors are persuaded to change votes during deliberations, their impressions of the defendant and the evidence might differ from the impressions of jurors who did not change votes. In order to investigate this possibility, a series of 2 x 2 analyses of variance were performed on memory, evidence ratings and defendant ratings; in which the independent variables were (1) initial verdict preference (guilty or not guilty) and (2) final verdict preference (guilty or not guilty). This analysis allowed comparison among four different groups, (a) jurors whose initial and final votes were guilty, (b) jurors whose initial and final votes were not guilty, (c) jurors who shifted from guilty to not guilty, and (d) jurors who shifted from not guilty to guilty. Of primary interest were interactions between initial and final votes, since an interaction would indicate that jurors differed as a function of changes that took place during the deliberation process.

Table 2-8 presents the initial and final vote distributions. Initially, 38% of all jurors voted to convict, and the final conviction rate was 28%, so jurors tended to shift in the direction of innocence. This is an example of the well documented majority persuasion effect that has been demonstrated in jury decisions (Penrod & Hastie, 1979, 1980; Stasser, Kerr, & Bray, 1982) as well as other small groups (Allen, 1965, 1975; Asch, 1951).

On post-deliberation measures, there were several main effects for initial and final votes, which will not be reported since they repeat analyses reported previously. There were only a few initial and final vote interactions, the means for these interactions are presented in Table 2-9. There were no interactions for memory measures.

For overall defense evidence ratings, there was a marginally significant initial x final interaction,  $F(1,585) = 3.26$ ,  $p = .07$ . Jurors who voted not guilty on both initial and final votes rated the defense evidence as stronger ( $M = 6.14$ ) than jurors who shifted from guilty to not guilty ( $M = 4.98$ ), whereas jurors whose final vote was guilty did not differ as a function of whether they had or had not changed their votes. This suggests that jurors who changed from guilty to not guilty were conforming to group pressure, but did not change their private impressions of defense evidence strength.

A complementary finding was obtained for defendant criminality ratings,

Table 2-8. Initial by Final Vote Distribution

	FINAL		
	Guilty	Not Guilty	
<u>INITIAL</u>			
Guilty	114 (19%)	114 (19%)	228 (38%)
Not Guilty	52 (9%)	320 (53%)	372 (62%)
n =	166 (28%)	434 (72%)	

Table 2-9. Initial x Final Vote Interactions on Post-Deliberation Data

#### 1. Defense evidence overall

<u>Initial</u>	<u>Final</u>	
	Guilty	Not Guilty
Guilty	3.71	4.98
Not Guilty	4.15	6.14

#### 2. Defendant criminality ratings

<u>Initial</u>	<u>Final</u>	
	Guilty	Not Guilty
Guilty	29.47	27.82
Not Guilty	28.30	23.81

in which the initial x final vote interactions was significant,  $F(1,550) = 4.61, p < .03$ . Jurors whose initial and final votes were not guilty rated the defendant more favorably ( $M = 23.81$ ) than jurors who changed from guilty to not guilty, suggesting that those who changed votes retained their private impression of defendant criminality. Again, jurors whose final vote was guilty did not differ as a function of whether they had changed votes.

A final interaction sheds additional light on the vote changing process-- this interaction was obtained on the verdict certainty ratings made for initial preferences prior to deliberations. The means for this interaction are presented in Table 2-10. Jurors who shifted from not guilty to guilty were initially less certain of their votes from the other three groups. This uncertainty may have led them to change their impressions, not only in terms of verdict, but for defendant and evidence ratings as well. This could have been the result of informational influence that took place during discussion (Deutsch & Gerard, 1955). Jurors who changed from guilt to innocence were initially as certain as those jurors who did not change votes. They may have changed votes simply as a result of majority pressure (since the majority did initially favor innocence), without changing their impressions of the defendant or evidence.

#### Post-Deliberation Memory Results

Analyses of post-deliberation questionnaire responses were performed on deliberating jurors' responses only, since deliberating and non-deliberating groups were no longer equivalent at the time they completed the questionnaire. In addition, individual subjects were no longer independent, since they had deliberated together in groups, so all analyses employed groups (nested within conditions) as the error term for significant tests.

Recognition task. The recognition task for joined conditions consisted of 16 multiple-choice items containing 4 correct and 4 incorrect items about the target offense, and 8 "intrusion items" that were facts from the two non-target offenses. Subjects were asked to choose as many or as few of these items as they thought were contained in the target charge testimony. Of course, subjects in the control group had not been exposed to any non-target testimony. They were given the recognition task containing the same correct and error items, along with intrusion items from all joined conditions (which, with few exceptions, were different for each condition), thus subjects in the control group had almost six times as many opportunities to make an "intrusion" error (a total of 45 possible intrusions). Therefore, the control-joined comparison represents a conservative test of the confusion hypothesis.

Overall, subjects were 91% accurate on the correct items, and made 10% factual errors on the incorrect items. There were no joinder effects on these measures. Our primary concern was with intrusion errors. There was a significant joinder effect on the number of intrusions,  $E(1, 90) = 15.40, p < .001, ES = .37$ , with a mean of .32 intrusions in the control group and .87 intrusions in experimental groups. There was a main effect for charge similarity,  $E(2, 90) = 4.76, p < .02$ , which indicated that subjects made more intrusions as charge similarity increased, with means of 1.10, .81, and .69 for identical, similar, and dissimilar charges respectively. There was a marginal Charge similarity x Evidence similarity interaction,  $F(2,90) = 3.59, p < .05, ES = .23$ , which was rather difficult to interpret. There

Table 2-10. Juror certainty in verdict as a function of initial and final vote.

INITIAL	FINAL		Initial x Final Interaction
	Guilty	Not Guilty	
Guilty	7.50	7.26	$F(1,583) = 4.09, p < .05$ .
Not Guilty	6.79	7.25	

1 = extremely uncertain, 9 = extremely certain

were fewer instructions in the similar charge-dissimilar evidence condition (M = .59) than in any other joined condition, so that this cell tended to "disrupt" the otherwise orderly pattern obtained for charge similarity. There was no effect for instructions, and instructions did not interact with any other variable. The recognition results as a whole indicate that joinder did promote confusion of evidence, but this confusion was not great relative to the total amount possible.

Evidence ratings. Subjects were asked to rate the overall strength of the evidence for prosecution and defense on 9-point scales from weak to strong, and also to rate the incrimination value of two individual evidence items for both prosecution and defense on 9-point scales from innocence to guilt. The responses to the two items were summed to produce four evidence ratings for each subject: (1) prosecution overall, (2) prosecution item sum, (3) defense overall, and (4) defense item sum. None of the analyses on these ratings yielded significant effects, although all means were in the predicted direction. Prosecution evidence overall was rated stronger (M = 4.23 vs. 3.70) and individual items were rated more incriminating (M = 11.73 vs. 10.98) in joined than single conditions; defense evidence was rated weaker (M = 5.26 vs. 5.66) and more incriminating (M = 8.44 vs. 8.03) in joined conditions. However, these differences were small and non-significant, and therefore offer little support for the hypothesis that joinder changes perceptions of evidence strength.

Defendant ratings. Subjects rated the defendant on eleven 9-point trait and behavior scales. Table 2-11 presents the means and standard deviations of eleven ratings, and the results of a factor analysis that yielded two factors. We have termed the first factor a "criminality-credibility" factor, and the second a "global evaluation" factor. For purposes of analysis, two factor scores were formed and subjected to the same analyses performed on the other dependent measures.

#### Relationships Among Variables

In order to investigate the process hypothesized to operate in joined trials, the relationships among the variables were examined using path analysis. Table 2-12 presents the zero-order correlations among the experimental manipulations, ratings of the defendant and the evidence, memory for evidence, and individual pre-deliberation verdicts. Based on our theoretical predictions, a causal model was devised to specify the hypothesized directional relationships among variables. Hierarchical regression analyses provided path coefficients representing the magnitude of these relationships. Figure 2 graphically depicts the results of the path analysis for the effects of the four manipulated variables and five mediating variables on verdict judgments. Dummy variable coding was employed for the manipulations of joinder (1 = joined, 0 = single), evidence similarity (1 = similar, 0 = dissimilar), and instructions (1 = present, 0 = absent). Charge similarity was scaled to reflect ratings from the manipulation check, resulting in codes of 1, 3, and 4 for dissimilar, similar, and identical charge conditions. Interactions were not coded, since there were no hypothesized interactions and virtually no interactions in the analyses previously reported. The mediating variables consisted of indicators of each of the three hypothesized mediating processes. Defendant criminality and evaluation scores provided measures of criminal inference, overall ratings of prosecution and defense evidence strength served

Table 2-11. Defendant ratings factor analysis

Statistics for each variable			Factor loadings	
Variable	Mean	SD	Factor 1	Factor 2
Sincere	5.09	1.96	.810	.000
Believable	5.10	2.10	.800	.000
Honest	5.07	1.99	.795	.000
Moral	5.06	1.54	.690	.000
Future crime	5.38	2.21	.690	.000
Likeable	4.85	1.69	.632	.000
Typical Criminal	4.89	1.95	.629	.000
Nervous	3.76	1.89	.000	.670
Good	4.90	1.50	.340	.657
Dangerous	3.80	2.19	.000	.626
Attractive	5.00	1.96	.000	.605

loadings less than .25 have been replaced by 0.

#### Correlation Matrix

Variable	Honest	D	L	G	S	A	B	N	M	F
Dangerous	.28									
Likeable	.42	.14								
Good	.36	.39	.26							
Sincere	.62	.17	.52	.29						
Attractive	.08	.13	.26	.26	.14					
Believable	.62	.14	.46	.25	.68	.09				
Nervous	.09	.22	.05	.24	.11	.20	.11			
Moral	.48	.21	.35	.28	.50	.02	.46	.08		
Future	.51	.21	.34	.34	.45	.09	.41	.07	.39	
Criminal	.43	.19	.33	.29	.35	.16	.39	.05	.37	.56

Table 2-12. Correlations between manipulations, ratings and verdicts

	Joinder	C	E	I	DC	DE	P	D	M	Mean	SD
Joinder										.90	.30
Charge similarity	.55									2.84	1.43
Evidence similarity	.40	.25								.59	.49
Instructions	.21	.12	.53							.29	.45
Defendant Criminality	.17	.17	.01	.00						25.90	7.07
Defendant Evaluation	.19	.13	.15	.07	.41					11.12	2.88
Prosecution Evidence	.07	.10	-.01	.03	.38	.13				4.24	2.50
Defense Evidence	-.07	-.09	-.03	-.04	-.34	-.18	-.48			5.37	2.21
Memory	.16	.18	.08	.04	-.04	-.03	.00	.03		.77	1.10
Verdict	.07	.03	-.02	.03	.33	.12	.55	-.39	-.01	.37	.48

as measures of accumulation of evidence, and the number of recognition intrusions was employed as the measure of confusion of evidence. Regression analysis revealed that these nine predictor variables accounted for 34% of the variance in the verdict data. The most important findings of the analysis are highlighted by the boldfaced lines in Figure 2, which represent all paths with coefficients of .10 or greater. Table 2-13 presents the path analysis results broken down into direct and indirect causal components.

The model was predicted on the hypothesis that joinder activates a criminal schema which affects jurors' verdicts both directly and indirectly through judgments of the defendant, evidence strength, and memory for evidence. With minor exceptions, the analysis strongly supported this prediction. Both joinder and charge similarity influenced memory directly; however, memory was unrelated to any other variables. Charge similarity was positively related to defendant criminality ratings in addition to memory, but bore little relationship to any other variables. Neither judges' instructions nor evidence similarity had direct or indirect effects on any of the variables.

Our primary concern was with the process whereby joinder influences jurors' decisions. Joinder had a small, positive, direct effect on verdict, while its influence on perceptions of the evidence was negligible. Joinder most strongly influenced perceptions of the defendant's criminality and global evaluations. Defendant criminality ratings influenced verdicts directly, and also strongly affected perceptions of the evidence, having a positive effect on prosecution ratings and a negative effect on defense ratings. However, global evaluations did not significantly influence verdicts or ratings of the evidence. Assessments of the evidence in turn affected verdicts, with strong positive effects for prosecution evidence and weaker, negative effects for defense evidence.

Since the path analysis is based on correlational data, the direction of the effects is not known; however, the results are consistent with the hypothesized pattern of causation. Further support for the hypothesis can be obtained from the decomposition of causal effects in Table 2-5. The strongest direct effects of verdicts were obtained for prosecution evidence; the strongest indirect effects came from defendant criminality ratings. Joinder exerted its strongest influence on defendant ratings. Thus, the results are consistent with a decisionmaking process whereby joinder leads to inferences about the defendant's criminality, which then influence verdicts both directly (perhaps based on judgments of representativeness) and indirectly (by influencing interpretation and accumulation of incoming evidence).

FIGURE 2: PATH MODEL OF REPRESENTATIVE JURORS' JUDGMENT PROCESS

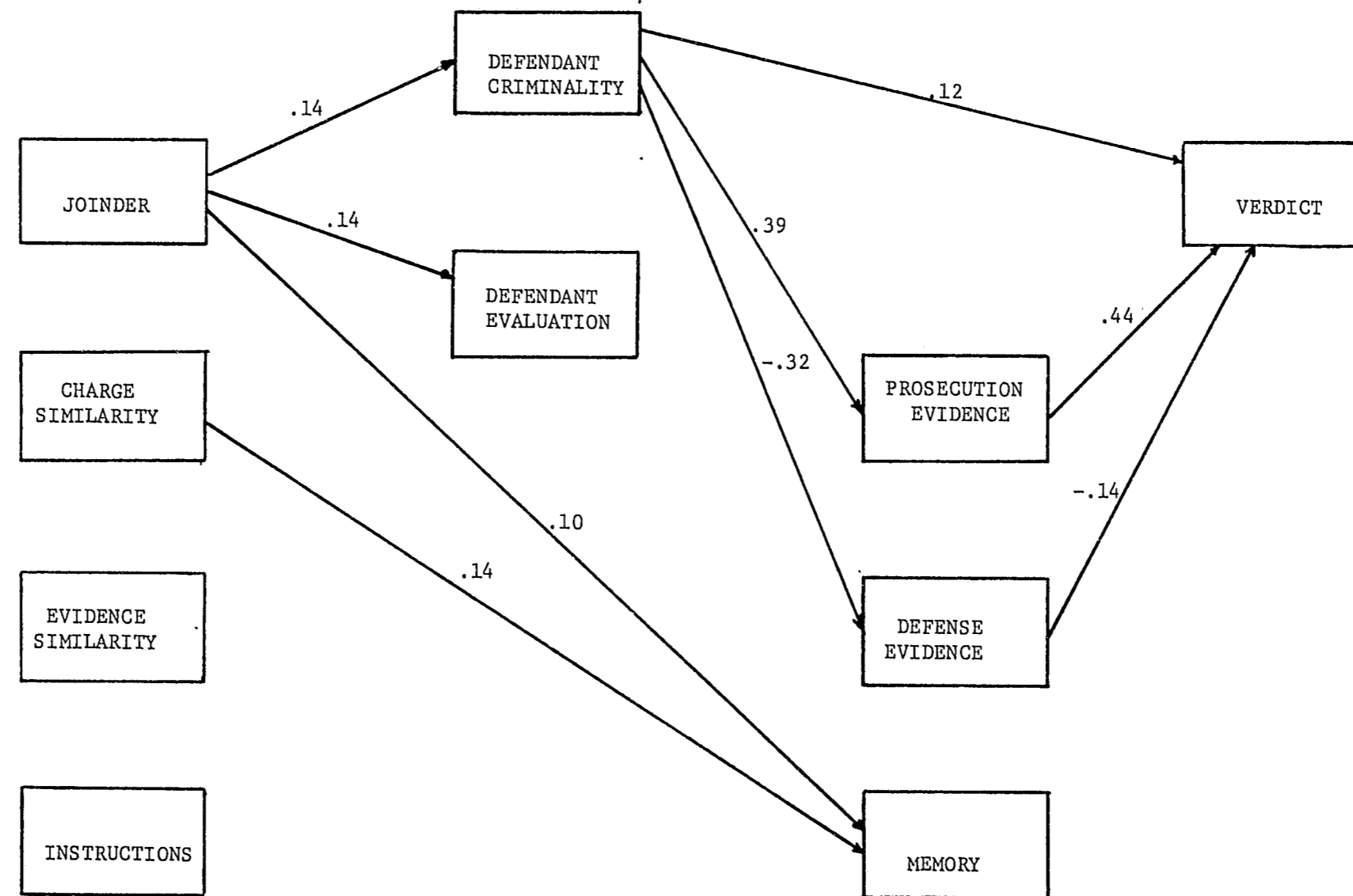


Table 2-13. Path analysis--Decomposition of causal effects on verdicts

	Total	Direct	Indirect
Joinder	.10	.06	.04
Charge similarity	-.01	-.07	.06
Evidence similarity	-.09	-.04	-.05
Instructions	.06	.03	.02
Defendant criminality	.34	.12	.21
Defendant evaluation	-.02	-.01	-.01
Prosecution Evidence	.44	.44	.00
Defense Evidence	-.14	-.14	.00
Memory	.00	.00	.00

CHAPTER 3. STUDY 3 (REPLICATION AND EXTENSION)

OBJECTIVES

The main study demonstrated significant joinder effects using representative jurors in a realistic trial setting including group deliberation. In order to achieve a high degree of external validity, the study was necessarily limited by time and financial constraints. The purpose of the third study was to replicate the conditions of the main study using undergraduates who did not deliberate, and to include a number of additional experimental conditions to produce a more complete design. The study employed a full factorial design manipulating charge similarity, evidence similarity, and judges' instructions. In addition to the target offense control group, single offense control groups for all non-target charges (the second and third charge in each joined condition) were included. The additional control groups served two purpose: (1) to extend the joinder effects obtained on the target charge to other charges, thus establishing the generality of the phenomenon, and (2) to investigate the magnitude of joinder effects as a function of the position of the charge in the joined sequence, as opposed to the first charge only. Study 3 also allows comparison between judgments made by undergraduates and the more representative jury pool population.

Method

Subjects. Subjects were 374 undergraduates at the University of Wisconsin who received course credit for participation. The sample was two-thirds female and one-third male, and subjects' mean age was 19 years.

Design. The design of the experiment is presented in Table 3-1. The study employed a 3 (charge similarity: identical, similar or dissimilar) x 2 (evidence similarity: similar or dissimilar) x 2 (instructions: present or absent) between-subjects design, where charge similarity, evidence similarity and instructions are defined as they were for Study 2. In addition, thirteen single-case control groups were run, consisting of the target control group (which was the first charge in all joined conditions), six single-case control groups corresponding to the second charge in each of the six combinations of charge and evidence similarity, and six single-case control groups corresponding to the third charge in each joined condition. Each of the 25 cells contained from 11 to 21 subjects, with a mean of 15 subjects per cell.

Stimulus materials. The study employed the same trial videotapes used in Study 2. In order to complete the factorial design, three additional joined tapes were employed consisting of the dissimilar evidence conditions in combination with the judges' instruction manipulation (Study 2 only included this manipulation in similar evidence conditions). In addition to the single target offense tape, twelve single case control tapes were prepared consisting of each non-target charge presented as a single trial. The content of the control tapes was identical to the content of the same offense presented in the joined trial. Each single trial lasted from 30 to 50 minutes, and each joined trial lasted from 1.5 to 2 hours.

Procedure. Subjects were randomly assigned to one of the 25 experimental conditions. Subjects viewed the trial videotape on a 19" television monitor. Following the trial, subjects individually completed a

Table 3-1. Experimental Design (Study 3)

Target Control Group	SINGLE CONDITIONS						JOINED CONDITIONS						EVIDENCE				
	Non-Target Control Groups						Identical			Similar					Dissimilar		
							CHARGES										
B <sub>1</sub> 0	B <sub>1</sub>	B <sub>1</sub>	b <sub>1</sub>	b' <sub>1</sub>	A <sub>1</sub>	R <sub>1</sub>	B <sub>1</sub>	B <sub>1</sub>	B <sub>1</sub> <sup>1</sup>	B <sub>1</sub>	b <sub>1</sub>	b' <sub>1</sub> <sup>3</sup>	B <sub>1</sub>	A <sub>1</sub>	R <sub>1</sub> <sup>5</sup>	Similar	NO INSTRUCTIONS
	B <sub>2</sub>	B <sub>3</sub>	b <sub>2</sub>	b' <sub>3</sub>	A <sub>2</sub>	R <sub>2</sub>	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub> <sup>2</sup>	B <sub>1</sub>	b <sub>2</sub>	b' <sub>3</sub> <sup>4</sup>	B <sub>1</sub>	A <sub>2</sub>	R <sub>3</sub> <sup>6</sup>	Dissimilar	
							B <sub>1</sub>	B <sub>1</sub>	B <sub>1</sub> <sup>7</sup>	B <sub>1</sub>	b <sub>1</sub>	b' <sub>1</sub> <sup>9</sup>	B <sub>1</sub>	A <sub>1</sub>	R <sub>1</sub> <sup>11</sup>	Similar	INSTRUCTIONS
							B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub> <sup>8</sup>	B <sub>1</sub>	b <sub>2</sub>	b' <sub>3</sub> <sup>10</sup>	B <sub>1</sub>	A <sub>2</sub>	R <sub>3</sub> <sup>12</sup>	Dissimilar	

Charge Codes

B = burglary (service station), b = burglary (residence), b' = burglary (business)  
 A = assault  
 R = robbery

Evidence Codes

1 = circumstantial evidence  
 2 = eyewitness identification  
 3 = other evidence (fingerprints, informant or stolen property)

questionnaire containing the same dependent measures used in Study 2: (1) verdict, certainty in verdict, probability of guilt; (2) background information, (3) trial ratings, (4) conviction criterion, (5) defendant ratings, (6) memory--evidence free recall, evidence recognition, recall of judges' instructions (instruction conditions only); (7) evidence strength ratings, (8) similarity ratings, questions pertaining to joinder (joined conditions only). A detailed description of the measures is provided in Study 2. All single trial sessions lasted approximately one hour altogether, and joined sessions lasted from 2 to 2.5 hours.

## RESULTS

Manipulation checks. Subjects in joined conditions rated the similarity of the charges and the evidence on 9-point scales. Responses to each question were analyzed in a 3 (charge similarity) x 2 (evidence similarity) x 2 (instructions) analysis of variance. For the charge similarity rating there was a main effect for charge similarity,  $F(2,171) = 15.06$ ,  $p < .001$ . The mean ratings for identical, similar and dissimilar charges were 3.33, 4.13 and 5.25, where a smaller number indicates greater similarity. There was also a marginally significant instruction effect for charge similarity ratings,  $F(1,171) = 3.34$ ,  $p = .07$ . Subjects who did not receive the instruction manipulation rated the charges as more similar ( $M = 4.0$ ) than subjects with instructions ( $M = 4.46$ ). For evidence similarity ratings, there was only a marginally significant effect for evidence similarity,  $F(1,171) = 2.79$ ,  $p = .10$ , which was rated as more similar in similar ( $M = 4.07$ ) than in dissimilar ( $M = 4.5$ ) evidence conditions. There was also a marginally significant charge similarity effect,  $F(2,171) = 2.79$ ,  $p = .06$ , and a marginal effect for instructions,  $F(1,171) = 3.26$ ,  $p = .07$ . The mean evidence similarity ratings in identical, similar and dissimilar charge conditions were 3.86, 4.44 and 4.47 respectively, and the mean ratings for no-instructions and instructions conditions were 4.07 and 4.50. The manipulation checks indicate that the charge similarity manipulation was successful, whereas the evidence similarity manipulation was weak. Therefore, the predicted effects of similarity should hold primarily for charge similarity, and not necessarily for evidence similarity.

Trial ratings. Mean overall ratings of interest, involvement, and trial realism were 4.45, 5.52, and 5.13 on 9-point scales. Therefore, subjects found their task to be a relatively engaging one.

As in study 2, a series of single degree-of-freedom contrasts was planned for analysis of target offense judgments involving the control group and various combinations of experimental groups. In addition, target offense judgments were analyzed in a 3 (charge similarity) x 2 (evidence similarity) x 2 (instructions) analysis of variance using the error term from the overall (13-cell) design including the target control group and the twelve experimental groups. Altogether the analyses employed 18 degrees of freedom, so the modified Bonferroni procedure for use with planned comparisons (Keppel, 1982) was used to set the acceptable significance level at .03. In addition to analyses on the target offense which were comparable to those performed for Study 2, supplemental analyses compared judgments on the second and third joined offenses with their single-case counterparts.

Verdict and guilt judgments. Table 3-2 presents the proportion of guilty verdicts obtained for the target offense in the single condition and the twelve joined conditions. The results revealed no significant differences between joined conditions and the control group, although in all but one cell the proportion of guilty verdicts in joined no-instructions conditions was higher than the proportion of guilty verdicts obtained in the control group. The C x E x I analysis of variance revealed a significant effect for instructions,  $F(1, 192) = 4.93$ ,  $p = .03$ , with fewer guilty verdicts with instructions ( $M = .31$ ) than without ( $M = .46$ ). This result is particularly interesting in light of the results of Study 2, in which instructions had no effect.

Certainty ratings. Again, unlike the previous study, there were joinder effects for certainty judgments. All single-joined comparisons were of at least marginal significance with the exception of the instructions versus control group contrast. Subjects in joined conditions expressed more certainty in their verdicts ( $M = 7.25$  overall) than subjects in the control group ( $M = 6.55$ ). There was also a main effect for instructions in the C x E x I analysis of variance,  $F(1,192) = 10.42$ ,  $p = .002$ . Subjects expressed less certainty with instructions ( $M = 6.88$ ) than without ( $M = 7.56$ ).

Probability of guilt. None of the analyses yielded significant results. The comparison between the identical charge condition and the control group was marginally significant,  $t(192) = 1.52$ ,  $p = .065$ , with higher guilt ratings for identical charges ( $M = 6.13$ ) than for the control group ( $M = 5.10$ ). The analysis of variance yielded a near-significant effect for instructions,  $F(1, 192) = 4.39$ ,  $p = .04$ , with lower guilt ratings with instructions ( $M = 4.90$ ) than without ( $M = 5.67$ ). In fact, the mean guilt rating was even lower for joined-instructions conditions than it was for the control group, although not significantly so.

Second charges. Table 3-3 compares the proportion of guilty verdicts obtained for the second charge in joined instructions and no-instructions conditions with the same offense judged as a single trial. Since the content of the charge was different in each experimental condition, analyses were performed on each offense (Row of Table 3-3) individually. With one exception (Row 4) there were more guilty verdicts in joined than single conditions, particularly without instructions, but only one of these differences was statistically significant. However, the analyses had low power due to small  $n$ 's of 13 to 21 subjects per cell. The mean joinder effect size across the six cases was .22, indicating that the effects were actually larger than those obtained in Study 2, for which the overall joinder effect size was .10.

For ratings of verdict certainty for the second charge. There was a tendency for subjects to be more certain in joined no-instructions conditions than in single conditions, although only one of these differences was statistically significant, and there was a tendency for certainty to be reduced in instructions conditions. Although weak, these results parallel the results obtained for certainty judgments on the first charge, where joinder increased certainty, and instructions decreased certainty. Analyses on probability of guilt ratings for the second charge revealed significantly higher guilt ratings for the identical charge-similar evidence condition ( $M =$



Table 3-2. Proportion of guilty verdicts--target charge (Study 3)

Control group	Charges			Evidence	
	Identical	Similar	Dissimilar		
.35 n=(20)	.53 (17)	.57 (21)	.20 (15)	similar	No Instruc- tions
	.54 (13)	.41 (17)	.47 (17)	dissimilar	
	.29 (14)	.41 (17)	.15 (13)	similar	Instruc- tions
	.20 (15)	.47 (15)	.27 (11)	dissimilar	

Table 3-3. Proportion of guilty verdicts--Charge in 2nd position

Control	Joined- No instructions	Joined- Instructions	Condition
.15a (13)	.35a (17)	.21a (14)	Identical charges Similar evidence
.20a (15)	.54a (13)	.27a (15)	Identical charges Dissimilar evidence
.19a (16)	.52b (21)	.35ab (17)	Similar charges Similar evidence
.43a (14)	.29a (17)	.27a (15)	Similar charges Dissimilar evidence
.07a (13)	.27a (15)	.38a (13)	Dissimilar charges Similar evidence
.07a (15)	.24a (17)	.09a (11)	Dissimilar charges Dissimilar evidence

For each row, means without common subscripts are significantly different at  $p < .05$

6.38) than its corresponding control group ( $M = 4.53$ ),  $t(40) = 2.15$ ,  $p = .04$ , as well as for the similar charge-similar evidence condition ( $M = 6.33$ ) in comparison to its control group ( $M = 4.31$ ),  $t(51) = 2.56$ ,  $p = .01$ . No other probability of guilt comparisons for Charge 2 reached significance.

**Third charges.** Table 3-4 presents the proportion of guilty verdicts for the third joined charges in comparison to their corresponding control groups. The results indicate that the third charge by itself was the weakest of the three offenses, particularly in identical and similar charge conditions, for which no guilty verdicts were obtained in control conditions depicted in Rows 1-3 of Table 3-4, and only a single guilty verdict is represented by the control group in Row 4. Joinder significantly increased the proportion of guilty verdicts in identical charge-similar evidence ( $M = .41$ ), identical charge-dissimilar evidence ( $M = .31$ ) and similar charge-similar evidence ( $M = .33$ ) conditions. The mean joinder effect size across the six cases was .28. Judges' instructions partially reduced the joinder effect for the identical charge-similar evidence condition, had no effect for the identical charge-dissimilar evidence condition, and were completely effective in the similar charge-similar evidence condition, reducing the amount of guilty verdicts to the baseline of zero. For dissimilar charge conditions, there were no differences between single, joined and instruction conditions. Joinder had no significant effects on certainty in verdict for any of the charges in the third position. However, joinder did affect probability of guilty judgments in a manner that paralleled the verdict results. The probability of guilt analyses on third charges revealed no differences between single, joined, and instructions groups.

The results of the analyses on Charge 2 and 3 offenses are not totally consistent with each other, probably because they are based on twelve different offenses. However, in conjunction with target offense judgments they present a fairly coherent picture, and establish the generality of joinder effects to a variety of cases. Joinder increased the probability of conviction on a given offense relative to the same offense tried alone, although this was primarily the case when charges were of the same nature (i.e., three burglaries) rather than three different offenses (burglary, assault, armed robbery). Judges' instructions reduced the conviction rate for joined charges, although the degree to which they were effective depended on the particular offense. Joinder had a tendency to increase subjects' certainty in their verdicts, and judges' instructions reduced verdict certainty.

**Recognition task.** Analysis of the recognition results was performed on the twelve experimental groups and the target control group only, since the task required subjects to choose which items were present in the target offense from among correct and incorrect target items and incorrect items from non-target offenses. Overall, subjects were 92% accurate on the correct items, and made 11% factual errors. Our primary concern was with intrusions or false recognitions of factors from other cases. There were 8 possible intrusions for experimental groups, and 45 possible intrusions for the control group. In all joined conditions, there were significantly more intrusions than there were in the control group. The  $C \times E \times I$  analysis of variance yielded a main effect for charge similarity,  $F(2,191) = 7.57$ ,  $p = .001$ , and a  $C \times E$  interaction,  $F(2,191) = 5.01$ ,  $p = .01$ . As predicted, the number of intrusions increased as a function of charge similarity, with means of .79, .90 and 1.56 in dissimilar, similar and identical charge conditions. The nature of the  $C \times E$  interaction was examined in an analysis of the simple effect of charge similarity for each level of evidence similarity. There was

Table 3-4. Proportion of guilty verdicts--Charge 3

Control	Joined- No instructions	Joined- Instructions	Condition
.00a (14)	.41b (17)	.21ab (14)	Identical charges Similar evidence
.00a (14)	.31b (13)	.31b (13)	Identical charges Dissimilar evidence
.00a (13)	.33b (21)	.00a (17)	Similar charges Similar evidence
.08a (13)	.35a (17)	.13a (15)	Similar charges Dissimilar evidence
.06a (16)	.13a (15)	.00a (13)	Dissimilar charges Similar evidence
.23a (13)	.35a (17)	.18a (11)	Dissimilar charges Dissimilar evidence

For each row, means without common subscripts are significantly different at  $p < .05$

no effect for charge similarity in dissimilar evidence conditions,  $F < 1$ , with means of 1.14, 1.26 and 1.29 for dissimilar, similar and identical charges. The charge similarity main effect was due to a very strong effect for charge similarity in similar charge conditions,  $F(2, 94) = 13.13$ ,  $p < .001$ , Effect size = .45, with means of .43, .61 and 1.81 for dissimilar, similar and identical charges. The results for the recognition task are consistent with those obtained in Study 2. Joinder led to confusion of evidence among charges, confusion increased as a function of charge similarity, but the amount of confusion was small relative to the total amount of possible confusion.

Recall. Free recall of evidence was scored to obtain four measures of recall for each charge: (1) total prosecution items, (2) total defense items, (3) proportion of prosecution items that were intrusions from other charges and (4) proportion of defense items that were intrusions. Analyses of intrusions were performed for joined conditions only, since recall intrusions were not possible in single-offense conditions. Analyses are reported for the target charge only, since the amount of information available for recall was different in each of the non-target charges. For total prosecution items recalled, there were no significant differences between experimental and control groups, and there were no main effects or interactions for the charge, evidence, and instruction manipulations. For defense evidence, there were significantly fewer items recalled in the similar charge condition ( $M = 1.94$ ) relative to the control group ( $M = 2.55$ ),  $t(90) = 2.38$ ,  $p = .02$ , and there were marginally fewer items recalled in the identical charge condition ( $M = 2.19$ ) than in the control group,  $t(190) = 1.89$ ,  $p = .06$ . An analysis of variance on the proportion of intrusions in joined groups only yielded no effects for any of the manipulations, and the mean proportion of intrusions overall was only 1%.

Taken as a whole, the results of evidence recall and recognition indicate that joinder did not produce much confusion among charges, although there was more confusion in joined trials than in a single trial.

Evidence ratings. Four measures of evidence strength were obtained for each subject for each charge: (1) prosecution evidence strength overall, (2) defense evidence strength overall, (3) prosecution item sum--the summed incriminating value (innocence to guilt) of two individual pieces of prosecution evidence, and (4) defense item sum (innocence to guilt) of two individual pieces of defense evidence. With the exception of the dissimilar charge condition, subjects in all joined conditions rated the overall prosecution evidence as stronger than subjects in the control group. The mean rating was 5.34 in joined conditions overall, and 4.05 in the control group on a scale from 1 (weak) to 9 (strong). The fact the joined-instruction mean (5.29) was marginally higher than the control group,  $t(191) = 2.06$ ,  $p = .04$  indicated that judges' instructions did not reduce subjects' perceptions of evidence strength, even though instructions did reduce their guilt verdicts. The C x E x I analysis of variance on the overall prosecution rating yielded a significant charge similarity effect,  $F(2, 191) = 3.94$ ,  $p = .02$ , with means of 4.62, 5.72 and 5.44 in dissimilar, similar and identical conditions. The effects obtained for the summed prosecution items were the same as those for the overall rating, except that there was no main effect for charge similarity. Subjects in all joined conditions, including those with instructions, rated specific items of prosecution evidence as more incriminating than subjects in the control group.

For the overall rating of defense evidence, there were no significant

differences between joined conditions and the control group, and there were no main effects or interactions for any of the manipulations. For the summed defense items, there was a marginally significant comparison between the similar charge condition and the control group,  $t(191) = 1.68$ ,  $p = .05$ , which indicated that subjects judging similar charges rated the evidence as more incriminating ( $M = 9.72$ ) than subjects in the control group ( $M = 8.20$ ). The C x E x I analysis of variance yielded a significant main effect for charge similarity,  $F(2, 191) = 8.02$ ,  $p < .001$ , with means of 7.71, 9.51 and 8.03 in dissimilar, similar and identical charge conditions. There was also an inexplicable C x E interaction,  $F(2, 191) = 3.69$ ,  $p = .025$ , for which the means in identical, similar and dissimilar charge conditions were 8.29, 10.03 and 7.00 with similar evidence, and 7.75, 8.87 and 8.84 with dissimilar evidence.

Defendant ratings. Subjects rated the defendant on the same eleven 9-point bipolar scales employed in Study 2. Table 3-5 presents the means, standard deviations and correlations for the defendant ratings. Table 48 also provides the results of a factor analysis on the ratings. The analysis yielded two factors, which were similar to those obtained with jury pool subjects. The first factor contained the items honest-dishonest, good-bad, moral-immoral, future crime likely-unlikely, believable-not believable, sincere-insincere, a typical criminal-not a typical criminal, dangerous-not dangerous, and likeable-dislikeable. This can be considered the criminality-credibility factor. The second factor had strong positive loadings on the items nervous-calm, attractive-unattractive, and a moderate loading on likeable-dislikeable. This can be considered to be a more global evaluation factor.

Defendant criminality and evaluation scores were created for each subject by summing their ratings for each item of the factor weighted by its factor loading. With the exception of the dissimilar charge condition, strong joinder effects were obtained for all experimental conditions; the defendant was rated less favorably in joined ( $M = 28.64$  overall) than single ( $M = 26.64$ ) trials. This was true for instructions ( $M = 28.18$ ) as well as no-instructions ( $M = 29.01$ ) groups. The C x E x I analysis of variance yielded a significant main effect for charge similarity,  $F(2, 355) = 4.72$ ,  $p = .01$ , and no other main effects or interactions. The defendant was rated less favorably when charges were identical ( $M = 29.10$ ) or similar ( $M = 30.48$ ) than when charges were dissimilar ( $M = 25.99$ ).

The same effects were obtained for the defendant evaluation factor, although they were not as strong. Again, ratings in all joined conditions were significantly higher than the control condition ratings, with the exception of the dissimilar charge condition. The joined-instruction mean (12.34) was marginally higher than the control mean (11.34),  $t(357) = 1.90$ ,  $p = .06$ , thus instructions were not totally effective in reducing joinder effects. There was a near-significant main effect for charge similarity on the defendant evaluation factor,  $F(2, 357) = 3.35$ ,  $p = .04$ . Again, subjects rated the defendant less favorably when charges were identical ( $M = 12.67$ ) or similar ( $M = 12.77$ ) than when they were dissimilar ( $M = 11.49$ ).

In conjunction with the defendant rating results of Study 2, these results offer the strongest and most consistent support for the prediction that joinder creates a negative impression of the defendant, which in turn

Table 3-5. Defendant Ratings Factor Analysis (Study 3)

Variable	Statistics for Each Variable		Factor Loadings	
	Mean	SD	Factor 1	Factor 2
Honest	4.44	2.02	.80	.00
Good	4.60	1.48	.77	.00
Moral	4.53	1.60	.74	.00
Future Crime Likely	4.27	1.96	.73	.00
Believable	4.68	2.07	.73	.32
Sincere	4.62	1.88	.72	.30
Typical Criminal	4.32	1.87	.69	.00
Dangerous	3.31	1.85	.68	.00
Likeable	4.82	1.64	.66	.42
Nervous	4.35	2.19	.00	.79
Attractive	5.68	1.51	.00	.61

loadings less than .25 have been replaced by zero

Correlation Matrix

	Honest	D	L	G	S	A	B	N	M	F
Dangerous	.44									
Likeable	.50	.38								
Good	.59	.55	.58							
Sincere	.64	.32	.56	.49						
Attractive	.27	.10	.39	.29	.26					
Believable	.68	.36	.52	.52	.72	.24				
Nervous	.12	.07	.20	.10	.14	.14	.22			
Moral	.55	.43	.54	.52	.54	.16	.56	.20		
Future Crime	.55	.36	.41	.48	.46	.24	.49	.06	.53	
Criminal	.49	.41	.51	.45	.45	.26	.42	.11	.48	.52

affects guilt judgments. In Study 2 there was a non-significant trend for the impression of the defendant to become less favorable as charge similarity increased, and the significant effects for charge similarity obtained in the present study strengthen this finding considerably. The results suggest that similar charges are more likely to create a coherent picture of the defendant as a prototypical criminal.

Joinder questions. Only subjects in joined conditions rated the degree to which the three charges together established a similar motive, intent, plan, identity, and criminal disposition on the part of the defendant. Responses to each question were analyzed using 3 (charge similarity) x 2 (evidence similarity) x 2 (instructions) analyses of variance. For the motive question, there was a main effect for charge similarity,  $F(2,172) = 17.16$ ,  $p < .001$ , a main effect for instructions,  $F(1,172) = 4.32$ ,  $p = .04$ , and no other effects. Subjects' ratings of motive similarity in dissimilar, similar and identical charge conditions were 4.49, 5.14 and 6.64 respectively, where a higher number indicates greater similarity. Subjects also rated motive to be less similar with instructions ( $M = 5.13$ ) than without ( $M = 5.56$ ). For the intent ratings there was a main effect for charge similarity,  $F(2, 172) = 46.31$ ,  $p < .001$ , a main effect for instructions,  $F(1,172) = 5.88$ ,  $p = .02$ , and a C x I interaction,  $F(2,172) = 3.67$ ,  $p = .03$ . Again, similar intent ratings increased as a function of charge similarity, with means of 4.20, 5.89 and 7.43 in dissimilar, similar and identical charge conditions. Subjects' intent ratings were lower with instructions ( $M = 5.59$ ) than without ( $M = 6.10$ ). An analysis of the simple effect of charge similarity for the C x I interaction revealed that there were significant charge similarity effects both with and without instructions, but the effect was stronger in the no-instructions condition,  $Epsilon = .63$ , than in the instructions condition,  $Epsilon = .53$ . For the question as to whether the three charges established a common plan, there was a single main effect for charge similarity,  $F(2, 172) = 26.85$ ,  $p < .001$ , with mean ratings of 3.47, 4.34 and 6.26 in dissimilar, similar and identical charge conditions.

In response to the question which asked whether the three charges established the identity of the criminal, there was a main effect for charge similarity,  $F(2,172) = 32.98$ ,  $p < .001$ , with mean ratings of 4.05, 4.89 and 6.98 in dissimilar, similar and identical charge conditions. There was also a C x E interaction for the identity question,  $F(2,172) = 3.53$ ,  $p = .03$ . Simple effects analyses revealed significant charge similarity effects for both levels of evidence similarity; however, this effect was stronger for similar ( $Epsilon = .62$ ) than for dissimilar ( $Epsilon = .35$ ) evidence conditions. For the question asking whether the three charges established a criminal disposition, there were main effects for charge similarity,  $F(2, 172) = 3.93$ ,  $p = .02$ , evidence similarity,  $F(1,172) = 3.91$ ,  $p = .05$ , and instructions,  $F(1,172) = 4.35$ ,  $p = .04$ , and no interactions. The disposition ratings were 3.69, 4.26 and 4.69 in dissimilar, similar and identical charge conditions, and subjects made higher disposition ratings in dissimilar ( $M = 4.51$ ) than similar ( $M = 3.96$ ) evidence conditions. Subjects also made lower ratings of criminal disposition with instructions ( $M = 3.92$ ) than without ( $M = 4.48$ ), which suggests that they were responding to the judges' instruction that they should not use multiple charges as evidence against the defendant. However, on the defendant ratings reported earlier, which provided a less direct measure of inferences about the defendant, instructions did not have an effect, and subjects in all joined conditions rated the defendant less favorably than subjects in single conditions. This suggests that subjects were not aware that joinder was affecting their

inferences about the defendant.

The final questions asked of subjects in joined conditions were whether they thought joinder was fair, whether they thought the three charges that they judged should be joined, and whether joinder had affected their judgments. For the "fair" question, there was a marginally significant charge similarity x instructions interaction,  $F(2,166) = 2.35, p = .10$ . Subjects' judgments of fairness increased as a function of charge similarity in the no-instructions condition, with means of 3.42, 4.14 and 5.03 for dissimilar, similar and identical charges; whereas there were no differences between levels of charge similarity in the instructions condition, means = 4.33, 4.48 and 4.33. In response to the question of whether charges should be joined, there was a marginally significant main effect for charge similarity,  $F(2,166) = 2.45, p = .09$ , with means of 3.43, 3.73 and 4.5 for dissimilar, similar and identical charges. For the question asking whether joinder had affected their judgments, there was a main effect for instructions,  $F(1,166) = 8.70, p = .004$ . Subjects who received instructions thought joinder affected their decisions less ( $M = 4.57$ ) than subjects who did not receive instructions ( $M = 5.72$ ). And, to a certain degree, they were right. Finally, subjects in instructions conditions only were asked how easy it was to follow the judges' multiple charge instruction, and a charge similarity x evidence similarity analysis of variance revealed no main effects or interaction. The mean rating in response to this question was 4.08, so on the average subjects were slightly below the neutral point in their judged ability to follow the instructions.

Verdict effects. The results of Study 2 indicated that jurors who preferred guilty verdicts had a lower criterion for conviction than jurors whose verdicts were not guilty, but that jurors' certainty about their judgments did not differ as a function of their verdict preference. Data for certainty and criterion measures in Study 3 were analysed using 2 (verdict: guilty or not guilty) x 2 (charge similarity) x 2 (evidence similarity) x 2 (instructions) analyses of variance. Only main effects for verdict and interactions of verdict with other variables are reported here, since the other effects are provided in previously reported analyses.

For certainty ratings, there was a significant verdict x instructions interaction,  $F(1,136) = 7.15, p = .008$ . An analysis of the simple effect of instructions revealed that subjects who voted to convict did not differ as a function of instructions,  $F < 1$ , with mean ratings of 7.52 with instructions and 7.41 without. For subjects preferring acquittal, there was a significant main effect for instructions,  $F(1,111) = 13.30, p < .001$ . Subjects were less certain of their not guilty verdicts with instructions ( $M = 6.46$ ) than without ( $M = 7.63$ ). Earlier results revealed that judges' instructions influenced undergraduate subjects' verdicts in the direction of acquittal. The verdict x instruction interaction suggests that those subjects who were influenced were less confident about their choice.

For the conviction criterion measure, there was a significant main effect for verdict,  $F(1,136) = 18.48, p < .001$ . Subjects whose verdicts were guilty reported a lower criterion ( $M = 80.11$ ) than subjects whose verdicts were not guilty ( $M = 88.73$ ).

## Summary of Results

Joinder did not significantly increase the proportion of guilty verdicts, although the proportion of guilty verdicts was higher in joined trials without instructions than in the control group. Significant joinder effects were obtained on verdicts for the second and third joined offenses in conditions involving three burglaries, but not in conditions involving three different charges. Judges' instructions were effective in reducing the proportion of guilty verdicts obtained in joined trials. Joinder increased subjects' certainty in their verdicts, and instructions in turn reduced verdict certainty. Joinder and instructions affected probability of guilt ratings in a manner that paralleled the verdict results. None of the manipulations affected subjects' conviction criterion or presumption of innocence.

Support was obtained for each of the three processes postulated to operate in joined trials. Subjects in joined trials made more recognition intrusions than subjects in the control condition, indicating a certain amount (although not a lot) of confusion between charges, which increased as a function of charge similarity. Subjects in joined trials rated the evidence for the prosecution as stronger than subjects in the control condition, whereas joinder did not have much influence on ratings of evidence for the defense. This suggests that subjects in joined trials were accumulating evidence against the defendant. Judges' instructions did not affect assessments of evidence strength. Subjects in all joined conditions rated the defendant less favorably on dimensions of criminality and global evaluation than subjects in single conditions. Negative inferences about the defendant increased as a function of charge similarity, and judges' instructions did not affect ratings of the defendant.

In sum, the results of Study 3 replicated and extended the findings of Study 2 with one exception. As in Study 2, joinder led to higher guilt judgments, confusion of evidence, accumulation of evidence against the defendant, and negative inferences about the defendant. The strongest effects were obtained when the joined charges were "identical" or similar. Unlike the representative jurors in Study 2, undergraduates' verdict and guilt judgments were influenced by judges' instructions. However, instructions had no effect on memory, judgments of evidence strength, or ratings of the defendant. This suggests that similar processes were operating in both groups, although undergraduates' verdicts were directly susceptible to the influence of instructions independent of perceptions of the defendant and the evidence. Analyses reported below will shed additional light on this issue.

## Relationships Among Variables

Table 3-6 presents the correlations among the manipulations, defendant and evidence ratings, memory and verdicts for the target charge. Hierarchical regression analyses were performed to test the same causal model applied to jury pool subjects against the undergraduate data. Since the charge similarity manipulation check indicated equal spacing between the three levels of charge similarity, codes of 1, 2 and 3 were used for dissimilar, similar and identical charges. Dummy coding was employed for the manipulations of joinder, evidence similarity and instructions. Mediating variables included defendant criminality and defendant evaluation scores, prosecution and defense overall evidence ratings, and recognition intrusions, with defendant ratings

Table 3-6. Mean Certainty Ratings in Undergraduate and Jury Pool Subjects as a Function of Verdict and Instructions

Verdict	UNDERGRADUATES		REPRESENTATIVE JURORS	
	No Inst.	Instructions	No Inst.	Instructions
Not Guilty	7.60	6.53	7.25	7.39
Guilty	7.35	7.52	7.44	7.47

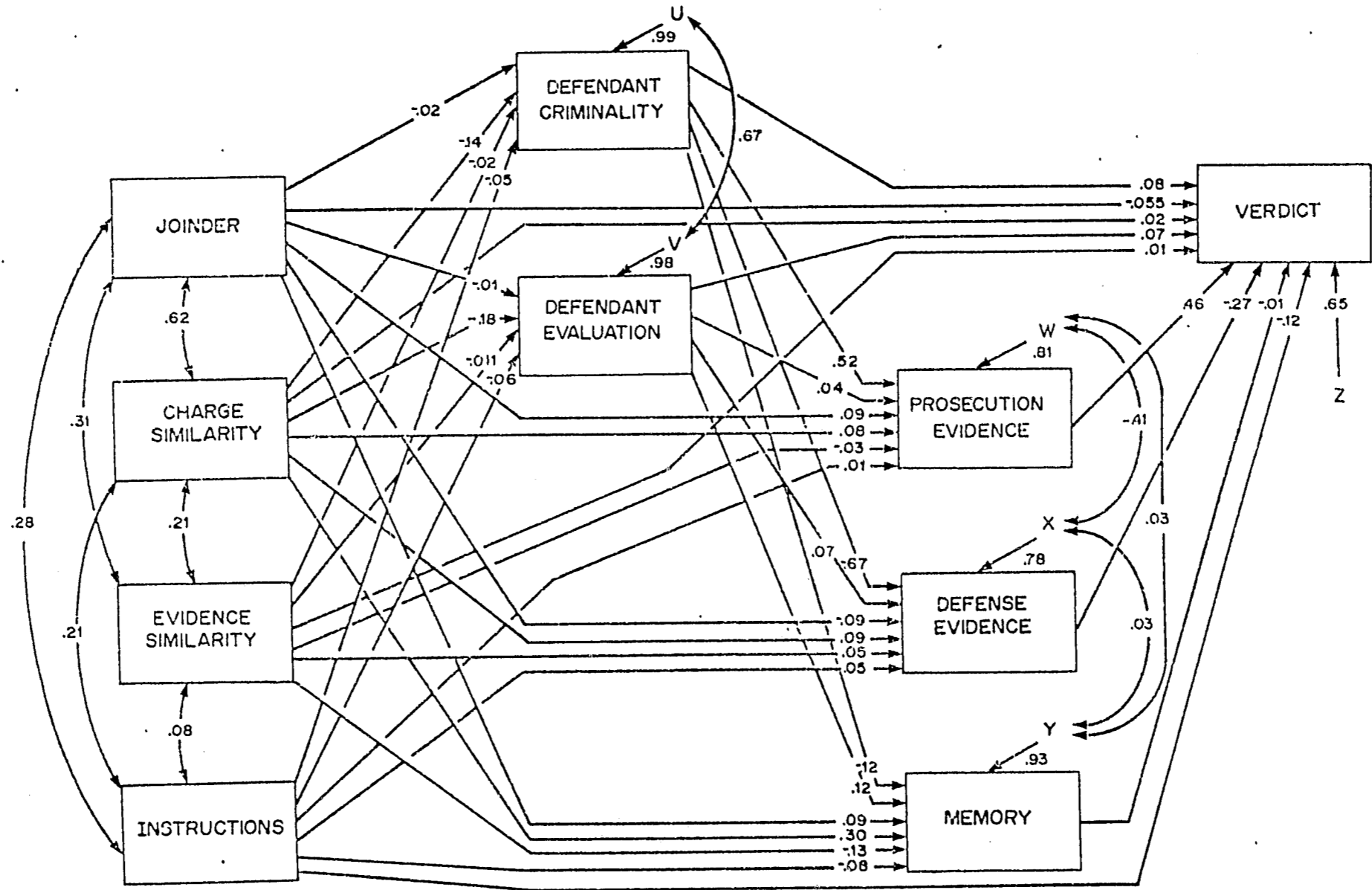
considered to be causally prior to evidence ratings and memory. Figure 3 presents a graphic representation of the model, with major path coefficients representing the magnitude of the effects between variables. Table 2-7 presents the decomposition of causal effects into direct and indirect components.

Due to the fact that the independent variables were positively correlated (particularly joinder and charge similarity,  $r = .62$ ), some of the joinder effects were obscured by the effects of other manipulations. The fact that the paths from joinder to defendant criminality and evaluation are small (and slightly negative) is not consistent with the prediction that joinder effects are mediated through inferences about the defendant. However, the zero-order correlations between joinder and defendant ratings in Table 2-13 are positive, although small (.05 and .08). The path coefficients between charge similarity and defendant criminality and evaluation ratings of .14 and .18 suggest that the charge similarity variable is suppressing the effects of joinder, due to the strong positive correlation between the two manipulations. Also, the correlations which served as input into the regression analyses were based on the target charge only--the twelve non-target control groups were not included, since they did not measure ratings of evidence, memory and verdicts for the target offense. In fact, the estimate for conviction rates in the target charge control group is based upon only 20 observations. However, the other twelve control groups ( $n = 157$  subjects altogether) did contain the same defendant ratings that were measured in all other conditions. Correlations between joinder and defendant factor scores from all single case control groups (coded as 0) and all experimental groups (coded 1) revealed correlations of .24 between joinder and defendant criminality, and .17 between joinder and defendant evaluation. These correlations indicate that joinder does lead to negative inferences about the defendant, although this effect is not represented in the path model, based upon data from the target control group only.

With one exception, the remaining results depicted in Figure 3 are consistent with the results obtained with jury pool subjects illustrated in Figure 2. Both joinder and charge similarity affected memory processes, but memory was unrelated to verdicts. Charge similarity led to inferences of defendant criminality and evaluation, and the correlations using all single and joined conditions indicate that joinder did also. Ratings of defendant criminality had a strong positive effect on prosecution evidence ratings (.52) and a strong negative effect on defense evidence ratings (-.67), whereas defendant evaluations were not strongly related to ratings of the evidence. Both defendant criminality and evaluations were directly related to verdicts, although these effects were small in magnitude (.08 and .07). The strongest effect on verdicts was obtained for prosecution evidence ratings (.46) followed by defense evidence ratings (-.27). Therefore, the results support the hypothesis that verdicts are a result of subjects' perceptions of the strength of the evidence, which are strongly influenced by judgments of the defendant's criminality.

The main difference between undergraduate and jury pool models concerned the relationship between instructions and verdicts. Instructions were unrelated to verdicts, or to any other variables, for jury pool subjects. In the undergraduate data, instructions had a direct negative impact on verdicts which was not mediated through any of the other variables. Therefore, it

Figure 3. Path model of undergraduate subjects' judgment process



appears that the effects of joinder on judgment processes in the two groups are quite similar, and that they differ primarily in terms of (1) the magnitude of the effects and (2) the susceptibility of verdicts (but not intervening processes) to the influence of judges' instructions.

#### A COMPARISON OF UNDERGRADUATES AND REPRESENTATIVE JURORS

The final results to be reported combined portions of the data from Studies 2 and 3 in a single analysis, including subject population as a factor. Only main effects and interactions involving the subject variable are reported. All data for the target offense were analyzed using 2 (subject population) x 3 (charge similarity) x 2 (evidence similarity) x 2 (instructions) analyses of variance. Data for selected variables were also analyzed using 2 (single versus joined) x 2 (subject population) analyses of variance to investigate the possibility of interactions of joinder with the subject variable.

Manipulation checks. For the charge similarity rating, there was a marginally significant main effect for subject population,  $F(1, 803) = 3.41, p = .065$ . Undergraduates rated the charges as more similar (4.18) than jury pool subjects (4.41). For evidence similarity ratings, there were no significant main effects or interactions involving the subject variable.

Verdict and guilt judgments. For individual verdict and probability of guilt judgments, there were no differences between subject populations, and subjects did not interact with any of the manipulations. It was noted earlier that a primary difference between the two studies was that undergraduates' verdicts were influenced by judges' instructions, whereas qualified jurors' verdicts were not. Therefore, one would expect an instructions x subject population interaction. This interaction did not even approach statistical significance,  $F(1, 815) = 1.76, p = .19$ , although the proportion of guilty verdicts in the undergraduate sample was lower with instructions (.30) than without (.46), whereas in the jury pool sample the proportion of guilty verdicts with and without instructions was virtually identical (.39 and .38).

The only effects for subject population with respect to verdict were obtained when the initial verdict preference was included as a factor in the analysis of ratings of certainty in verdict. There was a verdict x subject population x instructions interaction,  $F(1, 690) = 5.54, p = .02$ . The means involved in this interaction are provided in Table 2-6. In the undergraduate sample, there was an interaction between verdict and instructions; undergraduate subjects in instructions conditions whose verdicts were not guilty were less certain than subjects in all other conditions. Jury pool subjects had equally high levels of certainty in all conditions, which were also the same as undergraduates' certainty levels with the exception of the not guilty-instructions cell. This analysis confirms statistically confirms our earlier suggestion that while undergraduates' verdicts may be more easily influenced by instructions, they suffer from a lack of certainty as a result.

Also performed on the verdict, guilt and certainty ratings were 2 (joinder: single or joined) x 2 (subject population) analyses of variance, collapsed across similarity and instructions. Again, the only effect obtained was on the certainty variable, for which there was a marginally significant

joinder x population interaction,  $F(1, 630) = 3.43, p = .07$ . Undergraduate subjects expressed more certainty in their verdicts in joined ( $M = 7.50$ ) than single ( $M = 6.55$ ) conditions, whereas jury pool subjects did not ( $M = 7.27$  versus 7.20).

Conviction criterion. For the conviction criterion measure, there were no main effects or interactions of subject population with the similarity or instruction manipulations. However, when individual verdict preference was included as a factor, there was a verdict x population interaction on the criterion variable,  $F(1, 693) = 4.03, p = .045$ . The mean criterion ratings for undergraduates with guilty and not guilty verdicts respectively were 80.46 and 88.98, and for jury pool subjects the ratings were 83.00 and 85.87. This suggests that undergraduate verdicts were more sensitive to reasonable doubt criteria than jury pool verdicts. However, this result could also indicate that the deliberation process affected jury pool subjects' criterion ratings after they made their individual judgments, since the videotaped deliberations indicate that groups spent considerable time discussing the reasonable doubt standard.

Memory. There were no differences between subject populations on the evidence recognition task, free recall of evidence, or recall of judges' instructions, and subject population did not interact with any other variable.

Evidence ratings. For the rating of prosecution evidence overall, there was a main effect for subject population,  $F(2, 790) = 6.49, p = .002$ . Undergraduates rated the prosecution evidence as stronger ( $M = 5.27$ ) than jury pool subjects ( $M = 4.25$ ). There was also a main effect of subject population on the sum of individual prosecution items,  $F(1, 790) = 4.64, p = .03$ , which was rated as more incriminating by undergraduates ( $M = 12.42$ ) than by jury pool subjects ( $M = 11.74$ ). For defense evidence ratings, there were no differences between the two groups, either on overall ratings or individual items.

Defendant ratings. Since factor analyses on the defendant ratings yielded slightly different factors for jury pool subjects and undergraduates, a defendant rating score consisting of the sum of the eleven items was used for comparison purposes. Analysis of this measure revealed a marginally significant charge similarity x subject population interaction,  $F(2, 795) = 2.75, p = .065$ . This interaction was due to a charge similarity effect for undergraduates, with means of 49.52, 55.47 and 54.02 in dissimilar, similar and identical charge conditions; but not for jury pool subjects, with means of 53.22, 53.39 and 55.01.

Joinder questions. For the ratings of whether the three charges established a similar motive, intent, plan, identity and disposition, the only effect for subject population was a charge similarity x population interaction,  $F(2, 803) = 3.16, p = .04$ , for the motive question. As the analyses reported previously indicated, there were charge similarity effects for both groups, although the nature of these effects varied somewhat between populations. The mean motive ratings for identical, similar and dissimilar charges were 6.66, 5.07 and 4.42 for undergraduates, and 6.55, 5.52 and 3.86 for jury pool subjects. There were no differences between groups on ratings of whether joining charges was fair or whether charges should be joined. There was a main effect for subject population on ratings of the degree to which



joinder had affected subjects' judgments,  $F(1, 802) = 10.91, p = .001$ . Undergraduates thought joinder had affected their decisions more ( $M = 5.16$ ) than jury pool subjects ( $M = 4.57$ ).

Trial ratings. Undergraduate and jury pool subjects did not differ in their ratings of interest in the trial. Jury pool subjects reported being significantly more involved in their task ( $M = 6.24$ ) than undergraduates ( $M = 5.52$ ),  $t(1091) = 5.05, p < .001$ . This is likely due to participation in deliberation. On ratings of realism, undergraduates rated the trial as more realistic ( $M = 5.13$ ) than jury pool subjects ( $M = 4.68$ ),  $t(1091) = 3.06, p = .002$ .

#### Summary of Subject Population Comparisons.

In summary, there were surprisingly few differences in the judgment processes of undergraduates and representative jurors. The two groups did not differ statistically in terms of verdict and guilt judgments, although separate analyses of the two groups indicated that undergraduates' verdicts were affected by judges' instructions, whereas representative jurors were not. In jury pool joinder led to an increase in conviction rates. Undergraduates expressed less certainty if their verdicts were not guilty, and also expressed more certainty in joined than single trials. Representative jurors' certainty did not differ as a function of any of the manipulations. Undergraduate subjects' conviction criterion varied more than the criterion of jury pool subjects as a function of their verdicts. The two groups did not differ in their sentencing judgments.

There were no differences between subject populations on any of the memory measures. Undergraduates rated evidence for the prosecution, but not the defense, as stronger than jury pool subjects. Charge similarity affected undergraduates' ratings of the defendant to a greater degree than it affected representative jurors' ratings. The two groups did not differ in their perceptions of joinder, although undergraduates thought joinder had affected their judgments more than jury pool subjects. Finally, the two groups differed somewhat in their ratings of involvement and trial realism, although both groups' ratings were above average.

## CHAPTER 4: DISCUSSION

The present research has examined a number of issues concerning juror inferencing and judgment processes in multiple-offense trials. In this section, the main findings are discussed in terms of their theoretical and applied significance. First, a brief summary of the results of the two main studies is provided, and the results are assessed in light of empirical research on joinder. Second, the results as a whole are discussed in terms of the cognitive and social psychological approaches to juror decision making described in the introduction. Third, the methodological implications of the findings are noted, and finally, the legal implications of the research are discussed.

Study 2 was the main study of the research. The results of Study 2 are succinctly summarized in Figure 2. The results indicate that joining multiple charges in a realistic trial situation increases the proportion of individual guilty verdicts obtained on a particular (target) charge relative to the same charge tried by itself. The effects of the manipulations of charge and evidence similarity were relatively subtle compared to the effect of joinder of any sort. Convictions increased regardless of similarity, although there were more guilty verdicts when evidence was dissimilar than when it was similar. A very strong set of judges' instructions had no effect on verdict judgments whatsoever. Overall, the results are consistent with previous research using much less realistic methods. Other researchers find that joinder increases conviction rates and instructions do not significantly reduce convictions (Bordens & Horowitz, 1983; Greene & Loftus, 1981; Horowitz et al., 1980; Kerr & Sawyers, 1979; Tanford & Penrod, 1982). In the present study, joinder also increased the number of guilty and hung group verdicts, relative to the control group. This indicates that the biases induced in jurors prior to deliberation persist through deliberations and affect the final outcome.

Support was obtained for each of the three processes hypothesized to operate in a joined trial. Joinder led to a certain amount of confusion of evidence among charges, particularly when charges were similar, but memory was not related to individual verdicts. Joinder also led to an accumulation of evidence, as measured by ratings of evidence strength, particularly when charges were similar or evidence was dissimilar. Subjects in joined trials rated evidence for the prosecution as stronger than subjects in a single trial, and to a lesser degree rated the evidence for the defense as weaker. This suggests that subjects primarily accumulate evidence against the defendant. Ratings of evidence strength were strongly related to verdicts, more so for prosecution than for defense evidence. Joinder also led to negative inferences about the defendant on dimensions of criminality and global evaluation, and these ratings were significantly related to verdicts. Defendant criminality, but not general evaluation, was strongly related to judgments of evidence strength.

In terms of the three processes postulated to mediate joinder effects, the results are generally consistent with previous research using less realistic stimulus materials. Bordens and Horowitz (1983) and Tanford and Penrod (1982) both found that joinder led to confusion of evidence, although Tanford and Penrod found that confusion was not related to guilt judgments, whereas Bordens and Horowitz found that it was. Greene and Loftus (1981) and

Tanford and Penrod (1982) found that joinder led to negative inferences about the defendant, and Tanford and Penrod found that these inferences were strongly related to guilt judgments. Tanford and Penrod found that joinder led to higher ratings of evidence strength, although Bordens and Horowitz found that joinder did not affect ratings of thoughts against the defendant, which could be considered an indirect measure of evidence strength. However, both Bordens and Horowitz and Tanford and Penrod found that ratings were strongly related to verdict and guilt judgments.

The relationships among the variables in Study 2 were integrated into a causal model of judgment processes in joined trials, which is depicted in Figure 2. In the proposed model, joinder leads to negative inferences about the defendant's criminal character. These inferences affect verdicts both directly and indirectly through judgment of evidence strength, which in turn strongly affect verdicts. The theoretical and practical implications of this process are addressed below.

Study 3 replicated the main findings of Study 2 using undergraduates who did not deliberate, and extended the generality of joinder effects to additional, non-target offenses. As in Study 2, joinder increased the proportion of guilty verdicts relative to single-offense control groups primarily when the joined charges were the same as opposed to different crimes. In contrast to the Study 2 findings of a weak evidence similarity effect, evidence similarity did not affect verdicts, or any other judgments, probably because the evidence similarity manipulation was not successful for undergraduate subjects. Convictions increased as a function of charge similarity.

The magnitude of joinder effects on verdicts was influenced by the position of the charge in the joined sequence, with stronger effects obtained for later charges. The only consistently significant joinder effects were obtained for charges in the third position, although the magnitude of effects on the first two charges was similar to those obtained in Study 2, and the lack of significance may have been due in part to lower statistical power in Study 2. The position effects run counter to the findings of Bordens and Horowitz (1983) and Horowitz et al. (1980) who found that joinder primarily increased convictions on the first, but not the second of two joined charges.

Unlike the results of Study 2, judges' instructions significantly reduced the proportion of guilty verdicts in Study 3. This result is interesting not only because previous research found that instructions did not reduce joinder effects (Greene & Loftus, 1981; Tanford & Penrod, 1982), but also because the existing literature on limiting instructions in other domains indicates that they tend to be ineffective (Lind, 1982). Instructions also affected undergraduates' certainty in their verdicts. Compared to subjects who voted to convict and subjects who voted to acquit but did not receive instructions, subjects who voted to acquit following instructions were less certain about their verdicts. This suggests that instructions are effective at the expense of a loss of certainty in undergraduate subjects.

As in Study 2, Study 3 indicates that joinder leads to some confusion of evidence, perceptions of stronger prosecution evidence, and negative inferences about the defendant. Again, however, confusion was unrelated to verdicts, whereas evidence and defendant ratings were related to verdicts, as

well as to each other. The path model based on the data from Study 3 (Figure 3) indicates that processes very similar to those obtained with representative jurors were operating in undergraduate subjects. The main difference was that instructions had a direct influence on undergraduates' verdicts, without changing any of the intervening processes. A statistical comparison of all data from undergraduates and representative jurors revealed few differences between the two groups.

#### THEORETICAL IMPLICATIONS

In this section, the main findings of the research are interpreted in terms of the theoretical approaches to juror decision making that were outlined in the introduction. The cognitive and social implications of the results are discussed with respect to each of the three processes hypothesized to operate in joined trials: confusion, accumulation and criminal inference. The path diagrams in Figures 2 and 3 serve as reference points for this discussion.

#### Confusion

The results indicate that joinder led to a certain amount of confusion among charges on a recognition task, particularly when charges were similar. A strictly cognitive explanation for this result is that recognition intrusions were a result of interference effects in long term memory (Postman & Underwood, 1973). However, the fact that confusion increased as a function of charge similarity suggests a more social psychological explanation for these findings. Similar charges are more easily encoded into a single, coherent representation of the trial than are dissimilar charges. Therefore, it is likely that specific evidence items were recalled in relation to the overall schema, rather than for individual charges. This line of reasoning is supported by research conducted by Hastie and Kumar (1979), which indicated that subjects were more likely to recall schema-incongruent than schema-congruent information. Evidence from dissimilar charges should be less congruent with the overall schema than evidence from similar charges. Further support for a schema-based explanation of the memory results is indicated in the study by Sulim and Dooling (1974), which demonstrated that subjects were more likely to make memory intrusions for a passage that was high in schema-relatedness than one that was not.

Although joinder led to a certain amount of confusion of evidence between charges, confusion was unrelated to verdicts. This finding is consistent with research using other impression formation tasks, which indicates that memory for specific items of information is not strongly related to the overall impression of a stimulus (Anderson & Hubert, 1963; Dreben, Fiske & Hastie, 1979; Risky, 1979). The explanation given for these findings is that once the information has been integrated into an abstract representation of the stimulus, the overall impression is independent of the representation of specific items in memory (Dreben et al., 1979). This finding was obtained with "rich behavioral stimuli" (i.e. paragraphs, Dreben et al., 1979, p. 1764), as well as using trait adjectives (Anderson & Hubert, 1963; Risky, 1979). The present study yields a low correlation between memory for specific items and judgments with much richer stimuli than those previously used.

However, we would not want to argue that subjects' judgments were made

independent of their memory for any aspect of the trial. Research on the use of the availability heuristic (Tversky & Kahneman, 1973) indicates that people often make judgments on the basis of the most easily remembered information about a stimulus. For example, Reyes et al. (1980) found that manipulating the salience of arguments influenced subjects' judgments of a defendant's guilt. The lack of relationship between memory and verdicts in the present research was likely due to the fact that both recall and recognition tasks asked for memory of brief, discrete case facts, which were probably not the features most available to subjects when making guilt judgments.

On a free recall task, the only difference between joined and single conditions was that subjects in single conditions recalled more total evidence than subjects in joined conditions, both for prosecution and for defense. Joinder did not cause subjects to differentially recall more evidence against the defendant, and recall of evidence was unrelated to verdicts. The recall results underscore the implication of the recognition results that joinder-induced biases are not a result of memory processes.

#### Accumulation

The results indicate that subjects in joined trials rate the evidence for the prosecution as stronger than subjects in single trials, and to a lesser degree rate the evidence for the defense as weaker. The path analysis results further indicate that ratings of prosecution evidence are much more strongly related to verdicts than ratings of defense evidence. From an information integration perspective, this suggests that subjects assign more weight to evidence against the defendant (prosecution evidence) than evidence in favor of the defendant (defense evidence). This is consistent with research demonstrating that negative information is weighted more heavily than positive information in forming impressions (Anderson, 1965; Dreben et al., 1979; Fiske, 1980; Hamilton & Huffman, 1971; Hodges, 1974; Kanouse & Hanson, 1972). In addition, joinder clearly creates a negative overall impression of the defendant, which, if averaged in with judgments of both types of evidence, could make prosecution evidence appear stronger and defense evidence weaker.

If the negative impression of the defendant created by the multiple charge context is represented as a criminal schema, the same evidence rating results can be interpreted in a slightly different manner. Findings from studies reviewed in the introduction indicate that schemas guide the interpretation and organization of incoming information, and that information inconsistent with the schema is often distorted or ignored (Taylor & Crocker, 1981). The causal models in Figures 3 and 4 indicate strong relationships between judgments of defendant criminality and evidence ratings. This suggests that jurors distort the evidence to make it consistent with their criminal schemas, making prosecution evidence appear stronger and defense evidence appear weaker. The paths between evidence ratings and verdicts further suggest that jurors differentially use information to the extent that it is consistent with their schemas. That is, they seem to base their decisions to a greater degree on evidence against the defendant (prosecution evidence) which is consistent with a criminal schema, than evidence in favor of the defendant (defense evidence) which is more difficult to incorporate into a criminal schema.

#### Criminal Inference

The criminal inference hypothesis was tested by asking subjects to rate the defendant on various trait and behavioral characteristics. Factor analysis on these ratings yielded two factors, one representing the defendant's criminality and credibility, and the other representing more global evaluations. Analyses revealed that subjects in joined trials rate the defendant much less favorably on both dimensions than subjects in single trials. These inferences increase further as a function of charge similarity, particularly for undergraduate subjects.

From an attributional perspective, these results suggest that subjects are making inferences about the causes of the defendant's alleged criminal behavior based on the fact that he is charged with multiple crimes. The multiple charge situation provides information about behavior that is high in consistency, particularly when charges are similar, and thereby is likely to lead to an internal attribution. In terms of distinctiveness, the picture is not quite as clear. It could be argued that being charged with dissimilar crimes indicates behavior that is low in distinctiveness, since it is performed with respect to very different entities. If that were the case, the charge similarity results would not support an attributional interpretation. However, in all joined conditions, the crimes, even though they may have been similar in method, were committed against different victims, on different dates, and in different places, possibly indicating behavior that was low in distinctiveness and therefore more likely to lead to an internal attribution. Finally, in all joined conditions, the defendant's alleged criminal behavior could be considered low in consensus, thus the third component of an internal attribution was present.

Although the defendant rating results can be roughly characterized in attribution terms, they are more consistent with an interpretation that does not assume causal inferences are made in such a scientific manner. It was hypothesized that joinder creates an impression of the defendant as a prototypical criminal. The finding that defendant ratings became less favorable as charge similarity increased supports this hypothesis, since similar charges are more easily incorporated into a criminal schema than dissimilar charges. Defendant ratings were positively related to jurors' verdicts, and this relationship was stronger for ratings of defendant criminality than for global evaluations. Therefore, the defendant rating results suggest that joinder creates a criminal schema, which then influences verdicts to the extent that the defendant appears representative of a typical criminal.

#### Instructions

Judges' instructions had no effect on representative jurors' verdicts, and also did not influence jurors' memory, evidence ratings or defendant ratings. However, instructions did affect undergraduates' verdicts, although they did not influence any of the intervening processes. The jury pool findings are consistent with social psychological research on context effects (Asch, 1946), belief perseverance (Lord, Ross & Lepper, 1979; Ross, Lepper & Hubbard, 1975) and schema-based processing (Taylor & Crocker, 1981), all of which indicate that once impressions are formed, they are quite resistant to change. For the most part, the undergraduate results support these interpretations. Instructions did not affect undergraduates' ratings of the

defendant or the evidence, or their memory for evidence. Undergraduate subjects in joined conditions rated the defendant less favorably and the evidence as stronger than subjects in single conditions regardless of instructions. Instructions affected undergraduates' verdicts directly, with a concomitant loss of certainty in verdicts. This suggests that instructions do not affect undergraduates' impressions of the defendant, but that subjects are nevertheless likely to be influenced by instructions to change to a verdict they are not certain is correct.

#### Integration of Findings

In the introduction, we proposed a single explanation for the judgment process in joined trials which incorporates all three sources of prejudice: confusion, accumulation and criminal inference. The path model in Figure 2 provides a representation of this process, which is strongly supported by the results. The strongest direct effects of joinder were obtained for inferences of defendant criminality and evaluation, supporting the prediction that joinder creates a criminal schema, with accompanying negative evaluations of the defendant. The global evaluations, which perhaps represent the affective component of the impression, do not contribute further to the decision process. On the other hand, ratings of defendant criminality have a direct influence on verdicts, suggesting that verdicts are based in part on representativeness judgments. The influence of criminality is more strongly felt through perceptions of the evidence, which are distorted to make them appear consistent with the criminal schema; i.e. the prosecution evidence appears stronger, the defense evidence weaker in a joined trial. Perceptions of the evidence, in turn, strongly influence verdicts, particularly the evidence for the prosecution which has more diagnostic value with respect to impressions of the defendant's criminality. The impression of guilt that results from this process is quite resistant to the influence of judges' instructions.

#### METHODOLOGICAL IMPLICATIONS

The present research has implications for the issue of external validity with respect to jury simulation research methods (see Bray and Kerr, 1982, for a discussion of this issue). The results of laboratory research on joinder (Tanford & Penrod, 1982) were replicated using much more realistic stimulus materials, and for the most part the same effects were obtained in undergraduates and representative juror subjects. In the Tanford and Penrod (1982) study, as well as the present research, joinder led to increased conviction rates, and the same pattern of relationships among memory, defendant and evidence ratings, and verdicts was obtained. This indicates that the joinder phenomenon is a robust finding, and also demonstrates that effects can be obtained in the laboratory which have implications for more realistic settings. However, the magnitude of joinder effects decreased as trial realism increased. Tanford and Penrod (1982) obtained a mean joinder effect size of .25, the present Study 3 with undergraduates obtained a mean effect size of .21 (across all thirteen cases), and for the main study using representative jurors (Study 3) the overall joinder effect size was .10. This finding is consistent with a meta-analysis conducted by Linz et al. (1981) which indicates that as the realism of jury simulation methods increases, the magnitude of the effects of experimental manipulations decreases. The results also point to the need for replication of findings obtained in the laboratory

using more realistic materials, procedures, and subjects.

The present results revealed that judges' instructions influenced undergraduate jurors' verdicts but did not affect jury pool subjects. This suggests that the existing research concerning the effects of limiting instructions in a number of domains, most of which used undergraduate subjects, may not be generalizable to the juror population.

The results revealed few differences between the judgment processes of representative jurors who did and did not deliberate. Moreover, the biases induced in jurors' pre-deliberation verdicts were also found in group verdicts following deliberation. This result runs counter to the findings of Kaplan and Miller (1978), which suggest that deliberation can serve to correct juror biases. The present results suggest that biases obtained with individual jurors may be similar to those found in a courtroom situation that involves group deliberation. As a practical matter, to the extent that individual and group judgments are similar, the additional time and expense of including group deliberation in jury simulation studies may not always be necessary in order to obtain results that have applied significance. However, we should periodically check our findings to assure that they will hold for both individual and group judgments.

#### APPLIED IMPLICATIONS

Since the results of the main study were obtained using procedures that were high in external validity, they have clear applications to the courts. The study used representative juror subjects, realistic videotaped trials, and included group deliberation. In our discussion of applications we will focus on the main study (Study 2), since the undergraduate study (Study 3) used somewhat less realistic procedures. From an applied perspective, the most important dependent variable is verdict, along with the processes that mediate the effects of the manipulations on verdicts.

The results indicate that joinder increased the proportion of guilty verdicts on a particular target charge, relative to the same charge tried by itself. This effect was obtained at the level of both individual and group verdicts. At the individual level, 39% guilty votes were obtained in joined conditions overall, compared to 24% guilty verdicts in the control group, so joinder resulted in 15% more guilty verdicts than would otherwise be the case. Statistically, the magnitude of joinder effects was not large, with an effect size ( $r$ ) of .10 for the overall joinder effect. However, the results are of considerable practical significance, if the additional convictions are considered to be conviction errors. Although the absolute magnitude of joinder effects will depend upon numerous factors (type of crime, case strength, etc.), the present results indicate that joinder can substantially increase the chance that an innocent person will be convicted of a crime.

At the group level, joinder increased the number of guilty and hung jury verdicts, relative to the control group. Therefore, pre-deliberation biases persisted through group verdicts, and deliberation did not serve to correct these biases. This finding further emphasizes the applied significance of the results to a degree that would not be possible if deliberation procedures had not been used. The increase in hung juries in joined over severed trials has only tentative implications, since a time limit was placed on deliberations

which probably affected the hung jury rate. However, if it is the case that juries are hung more often when deliberating on joined charges, this suggests that some of the supposed expedience of trying multiple charges together may be offset by an increase in hung juries.

#### Guidelines for Joinder

From an applied perspective, the study had two main objectives: (1) to develop guidelines delineating situations in which joinder would and would not be prejudicial, and (2) to design a set of instructions that would effectively reduce prejudice resulting from joinder. With respect to the first goal, the results indicated that the effects of the charge and evidence similarity manipulations were quite subtle compared to the effects of joinder of any sort. Regardless of the experimental condition, there were more convictions on the target charge in the context of a joined trial than on the same charge tried alone. There was a tendency for jurors to convict more often in dissimilar evidence than similar evidence conditions. A likely explanation for this finding is in terms of the probative value of the evidence. Independent ratings of the evidence by a group of undergraduates indicated that evidence defined as "dissimilar" was rated higher than evidence defined as "similar" in terms of its credibility, value, and informativeness. Since ratings on these three measures were highly correlated, together they can be considered a measure of probative value. Although the probative value of evidence for the target charge should have remained the same in all conditions (since it was always the same evidence), subjects apparently used their perceptions of the evidence in non-target charges when making target charge judgments, and therefore convicted more often in dissimilar evidence conditions.

Since joinder significantly increased convictions in all experimental conditions, one possible guideline would be to (1) avoid joining charges at all. This solution can be compared to two legal criteria currently used as a basis for some joinder decisions. The "simple and distinct" test holds that charges can be joined if the evidence from each is simple enough that jurors will not confuse evidence between charges. The present results indicate that this solution is not likely to reduce prejudice. Subjects judging joined offenses did confuse evidence between charges, but confusion was unrelated to verdicts.

The law primarily allows for joinder of similar crimes, and charges are often joined if they can pass the "other crimes" test of admissibility. Rule 404 (b) of Federal Rules of Evidence specifies that evidence of other actions is relevant (what we shall term "legally relevant") under certain conditions:

Evidence of other crimes, wrongs, or acts is not admissible to prove the character of a person to show that he acted in conformity therewith. It may, however, be admissible for other purposes, such as proof of motive, opportunity, intent, preparation, plan, knowledge, identity, or absence of mistake or accident.

Jurors in the present study were asked the degree to which the three charges they judged established a similar motive, intent, a common plan, and the identity of the criminal--these are elements of similarity that are legally relevant from one charge to another. Subjects were also asked the

extent to which the three charges established a criminal disposition on the part of the defendant--evidence from other crimes is not legally admissible for this purpose. The results indicated that ratings of motive, intent, plan and identity increased significantly as a function of charge similarity as defined in the experimental manipulations, whereas disposition ratings did not. This suggests that jurors' assessments of relevance are similar to the assessments of legal professionals as embodied in Rule 404 (b). Therefore, if the law wants jurors to use evidence from other crimes only when it is legally relevant, a possible guideline for joining charges would be to (2) adopt extremely stringent standards for the application of rules such as Rule 404 (b).

The law allows joinder of similar charges which would fall into our operationally defined identical and similar charge categories. The charges defined as "identical" clearly met the other crimes requirement, whereas those defined as "similar" fell somewhere in the "grey area" where it was not clear whether they were legally relevant or not. The results indicated that jurors' verdicts were influenced equally in both conditions. However, increased convictions as a result of joined trials that fall into the grey area of legal relevance can be considered more prejudicial, and therefore more likely to be appealed, than convictions for joined charges which are clearly connected. Therefore, stringent guidelines could serve to eliminate joinder of charges that fall into the grey area.

#### Impact of Instructions

In addition to establishing guidelines for joinder decisions, a second applied goal of the research was to devise a set of judges' instructions that would alleviate prejudice caused by joinder. Although social psychological research as well as previous empirical work on instructions cast some doubt on whether it would be possible to develop effective instructions, we did want to give the traditional legal remedy for prejudice a fair test. Therefore, a strong set of instructions was devised, patterned after existing instructions yet longer and more complete. The instruction contained elements corresponding to each of the three legal theories of prejudice: confusion, accumulation and criminal inference.

The results of the main study indicated that instructions had no effect whatsoever for representative juror subjects. Viewed in light of other failures to develop effective instructions, the present results strongly indicate that the current legal remedy for prejudice resulting from joinder may simply not be adequate. In order for instructions to be effective, they would need to disrupt the processes that mediate the effects of joinder on verdicts. The causal model depicted in Figure 2 suggests that joinder effects are mediated through a criminal schema for the defendant. The portion of the instruction manipulation that addressed this process stated that the fact that the defendant was charged with more than one crime should not be used as evidence against him. If, as we have argued throughout, criminal inferences are not a byproduct of a rational, strictly cognitive process, then it is not surprising that simply instructing jurors not to make inferences did not work. It is not clear that any instructions could be effectively change these inferences. Perhaps if an alternative, competing schema could be activated, which would prevent jurors from using a criminal schema, the biased judgment process in joined trials could be undermined. If this could be accomplished,

and if our causal model is correctly formulated, then joinder would also not affect ratings of the evidence (which are presumably mediated through criminal inferences) and as a result, convictions would not increase. Only further research can establish whether viable instructions can be developed.

The applied significance of the present results extends beyond the issue of joinder of offenses to other analogous trial situations in which similar processes might operate. As noted in the introduction, there are other forms of joinder in addition to joinder of distinct crimes occurring at different times and places. A defendant may be charged with multiple crimes arising out of a single act (same transaction joinder), and more than one defendant can be tried in a single trial (joinder of defendants). In a related vein, jurors may be allowed to choose from among several verdicts alternatives with respect to a particular crime, and empirical research indicates that the order and seriousness of decision alternatives can affect the verdict reached (Kerr, 1978; McComas & Knoll, 1974; O'Brien et al., 1983; Vidmar, 1972). Although there is little empirical evidence concerning the social inferences processes involved in the above multiple charge situations, the present results suggest that these processes could be meaningfully studied by investigating factors that mediate the effects of the initial phenomena on the final outcome.

The present results have additional significance insofar as they suggest ways in which jurors might become biased as a function of other evidentiary and procedural factors. Central to the Rules of Evidence is the "balancing test," whereby prejudice is weighed against probative value to determine the admissibility of potentially prejudicial evidence. The relevance of evidence of other crimes to the joinder issue has already been discussed. Evidence of prior convictions might produce inferential biases similar to those found in joined trials. As with other crimes evidence, evidence of prior convictions is not admissible for the purpose of establishing a criminal disposition. However, evidence of prior convictions seems even more likely to be prejudicial than joinder, since the defendant has actually been convicted of previous offenses, rather than just being charged with more than one crime. Therefore, a process model similar to the one proposed for joinder effects might also apply to the effects of prior convictions.

Since the present results suggest that joinder effects are mediated through inferences about the defendant's character, the findings have implications for the prejudicial effects of character evidence, which can also be introduced for the purpose of attacking witness credibility. In fact, introducing damaging character evidence may be sufficient to create criminal inferences of the type induced by joinder, which would then affect assessments of evidence strength and therefore verdicts. As noted by Penrod and Borgida (1983), research indicates that character evidence is likely to be prejudicial. Therefore, if character evidence evokes biases similar to those found in joined trials, one way to investigate these processes would be to experimentally manipulate character evidence in order to "activate" various schemas about the defendant.

The effect of judges' instructions with respect to joinder has implications for the use of limiting instructions designed to alleviate the effects of other types of prejudicial evidence. A strong and complete set of instructions had no effect on representative jurors' verdicts, suggesting that traditional legal remedies for alleviating prejudice are not adequate. A

future direction with respect to instructions might be to move to non-traditional devices that attack the problem at its source rather than its outcome, by undermining the inferential processes that result in biased judgments.

The Rules of Evidence and Criminal Procedure are replete with intuitive psychological assumptions that are conducive to empirical testing. Empirical research has investigated the effects of various types of prejudicial evidence and trial procedures on subjects' decisions, and well as the effects of limiting instructions--these studies are reviewed in the introduction. For the most part these studies have been demonstrations of simple cause and effect relationships. They tell us that a particular kind of evidence either does or does not affect judgments, and that instructions either are or are not effective in removing these effects, but do not tell us how these effects operate. In order to provide adequate answers to the questions posed in the law, an understanding of the psychological mechanisms underlying these effects is needed. The present results suggest that by conducting theoretically-grounded research in a legally informed manner, using methods that are generalizable to the courts, it may be possible to develop guidelines that can assist the courts in making their decisions, and relieve the courts of the need to rely upon informed intuitions.

### References

- Allen, V. L. (1965). Situational factors in conformity. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 2). New York: Academic Press.
- Allen, V. L. (1975). Social support for nonconformity. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 8). New York: Academic Press.
- Anderson, N. H. (1965a). Averaging versus adding as a stimulus combination rule in impression formation. Journal of Experimental Social Psychology, 70, 394-400.
- Anderson, N. H. (1965b). Primacy effects in personality impression formation using a generalized order effect paradigm. Journal of Personality and Social Psychology, 2, 1-9.
- Anderson, N. H. (1966). Component ratings in impression formation. Psychonomic Science, 6, 279-280.
- Anderson, N. H. (1974). Cognitive algebra. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 7). New York: Academic Press.
- Anderson, N. H. (1978). Progress in cognitive algebra. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 11). New York: Academic Press.
- Anderson, N. H., & Hubert, S. (1963). Effects of concomitant verbal recall on order effects in personality impression formation. Journal of Verbal Learning and Verbal Behavior, 2, 379-391.
- Anderson, N. H., & Lampel, A. K. (1965). Effect of context on ratings of personality traits. Psychonomic Science, 3, 433-434.
- Asch, S. (1946). Forming impressions of personality. Journal of Abnormal and Social Psychology, 48, 258-290.
- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), Groups, leadership, and men. Pittsburgh: Carnegie Press.
- Bar-Hillel, M. (1973). On the subjective probability of compound events. Organizational Behavioral and Human Performance, 9, 396-406.
- Baron, S. A. (1977). A look at the Tennessee multiple offender and the joinder and severance of criminal offenses for trial. Memphis State University Law Review, 7, 457-473.
- Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. Cambridge: Cambridge University Press.
- Bordens, K. S., & Horowitz, I. A. (1983). Information processing in joined and severed trials. Journal of Applied Social Psychology, 13, 351-370.
- Bray, R. M., & Kerr, N. L. (1982). Methodological considerations in the study of psychology of the courtroom. In N. L. Kerr & R. M. Bray (Eds.), The psychology of the courtroom. New York: Academic Press.
- Bruton v. U.S. 391 U.S. 123 (1968).
- Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.
- Cleary, E. W. (Ed.). (1972). McCormick's handbook of the law of evidence (2nd Ed.). St. Paul, MN: West.
- Crocker, J. (1981). Judgment of covariation by social perceivers. Psychological Bulletin, 90, 272-292.
- Crocker, J., Hannah, D. B., & Weber, R. (1983). Person memory and causal attributions. Journal of Personality and Social Psychology, 44, 55-56.
- Devitt, E. J., & Blackmar, C. B. (1977). Federal Jury Instructions and Practice. St. Paul, MN: West.
- Dreben, E. K., Fiske, S. T., & Hastie, R. (1979). The independence of evaluative and item information: Behavior-based impression formation. Journal of Personality and Social Psychology, 37, 1758-1768.
- Drew v. United States. (1964). Federal Reporter (2d series), 331, 85-94.
- Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. Journal Personality and Social Psychology, 38, 889-906.
- Greene, E., & Loftus, E. (in press). When crimes are joined at trial: Institutionalized prejudice? Law and Human Behavior.
- Hamilton, D. L. (1981). Cognitive representations of persons. In E. T. Higgins, C. P. Herman & M. P. Zanna (Eds.), Social cognition: The Ontario symposium (Vol. 1). Hillsdale, NJ: Erlbaum.
- Hamilton, D. L., & Huffman, L. J. (1971). Generality of impression formation processes for evaluative and nonevaluative judgments. Journal of Personality and Social Psychology, 20, 200-207.
- Hastie, R. (1981). Schematic principles in human memory. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), Social cognition: The Ontario symposium. Hillsdale, NJ: Erlbaum.
- Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. Journal of Personality and Social Psychology, 37, 25-38.
- Hastie, R., Ostrom, T. M., Ebbesen, E. B., Wyer, R. S., Hamilton, D. L., & Carlston, D. E. (Eds.). (1980). Person memory: The cognitive basis of social perception. Hillsdale, NJ: Erlbaum.

- Higgins, E. T., Herman, C. P., & Zanna, M. P. (1981). Social cognition: The Ontario symposium (Vol. 1), Hillsdale, NJ: Erlbaum.
- Himmelfarb, S. (1975). What do you do when the control group doesn't fit into the factorial design? Psychological Bulletin, 82, 363-368.
- Hodges, B. F. (1974). Effect of valence on relative weighting in impression formation. Journal of Personality and Social Psychology, 30, 378-381.
- Holderman, L. (1977). Joinder of offenses in Nebraska: Judicial discretion v. fair and impartial trial. Nebraska Law Review, 56, 399-409.
- Horowitz, A., Bordens, K. S., & Feldman, M. S. (1980). A comparison of verdicts obtained in severed and joined criminal trials. Journal of Applied Social Psychology, 10, 444-456.
- Hovland, C., Janis, F., & Kelley, H. H. (1953). Communication and persuasion. New Haven, CT: Yale University Press.
- Kahneman, D., Slovic, S., & Tversky, A. (1982). Judgments under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.
- Kanouse, D. E., & Hanson, L. R. (1972). Negativity in evaluations. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins & B. Weiner (Eds.), Attribution: Perceiving the causes of behavior. Morristown, NJ: General Learning Press.
- Kaplan, M. F. (1971). Context effects in impression formation: The weighted average versus the meaning change formulation. Journal of Personality and Social Psychology, 19, 92-99.
- Kaplan, M. F. (1974). Context-induced shifts in personality trait evaluation: A comment on the evaluative halo effect and the meaning change interpretation. Psychological Bulletin, 81, 891-895.
- Kaplan, M. F. (1975). Evaluative judgments are based on evaluative information: Evidence against meaning change in evaluative context effects. Memory and Cognition, 3, 375-380.
- Kaplan, M. F., & Miller, L. E. (1978). Reducing the effects of juror bias. Journal of Personality and Social Psychology, 36, 1443-1455.
- Kassin, S. M., & Wrightsman, L. S. (1979). On the requirements of proof: The timing of judicial instructions and mock juror verdicts. Journal of Personality and Social Psychology, 37, 1877-1887.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), Nebraska symposium on motivation. Lincoln, NB: University of Nebraska Press.
- Kelley, H. H. (1972). Causal schemata and the attribution process. Morristown, NJ: General Learning Press.

- Keppel, G. (1982). Design and analysis: A researcher's handbook (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Kerr, N. L. (1978). Severity of prescribed penalty and mock jurors' verdicts. Journal of Personality and Social Psychology, 36, 1431-1442.
- Kerr, N. L., & Sawyers, G. W. (1979). Independence of multiple verdicts within a trial by mock jurors. Representative Research in Social Psychology, 10, 16-27.
- Lamm, H., & Myers, D. G. (1978). Group-induced polarization of attitudes and behavior. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 11). New York: Academic Press.
- Langer, E. J., & Abelson, R. P. (1974). A patient by any other name . . . : Clinician group difference in labeling bias. Journal of Consulting and Clinical Psychology, 42, 4-9.
- Lempert, R. O., & Saltzburg, S. A. (1983). A modern approach to evidence (2nd ed.). St. Paul, MN: West.
- Lilly, G. C. (1978). An introduction to the law of evidence. St. Paul, MN: West.
- Lind, E. A. (1982). The psychology of courtroom procedure. In N. L. Kerr & R. M. Bray (Eds.), The psychology of the courtroom. New York: Academic Press.
- Linz, D., Slack, A., Kaiser, K., & Penrod, S. (1981). Meta-analysis of defendant characteristic studies. Presented at the American Psychology-Law Society Association Meeting, Boston.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of Personality and Social Psychology, 37, 2098-2109.
- Luchins, A. S. (1957). Primacy-recency in impression formation. In C. I. Hovland (Ed.), The order of presentation in persuasion. New Haven, CT: Yale University Press.
- McComas, W. C., & Noll, M. E. (1974). Effects of seriousness of charge and punishment severity on the judgments of simulated jurors. Psychological Record, 24, 545-547.
- Nisbett, R. E., Borgida, E., Crandall, R., & Reed, H. (1976). Popular induction: Information is not always informative. In J. S. Carroll & J. W. Payne (Eds.), Cognition and social behavior. Hillsdale, NJ: Erlbaum.
- Nisbett, R., & Ross, L. (1980). Human inference: Strategies and shortcomings of social judgment. Englewood Cliffs, NJ: Prentice-Hall.
- Note. (1964-1965). Joint and single trials under Rules 8 and 14 of the Federal Rules of Criminal Procedure. Yale Law Review, 74, 553-566.



- O'Brien, M., Williams, K. D., & Greenberg, J. (1983). Serial position effects of verdict instructions on jury decision making. Presented at the Midwest Psychological Association Meeting, Chicago.
- Ostrom, T. M. (1977). Between-theory and within-theory conflict in explaining context effects in impression formation. Journal of Experimental Social Psychology, 13, 492-503.
- Penrod, S., & Borgida, E. (1983). Legal rules and lay inference. In L. Wheeler (Ed.), Review of personality and social psychology (Vol. 3). Beverly Hills, CA: Sage.
- Penrod, S., & Hastie, R. (1979). Models of jury decisionmaking: A critical review. Psychological Bulletin, 86, 462-492.
- Penrod, S., & Hastie, R. (1980). A computer simulation of jury decisionmaking. Psychological Review, 87, 133-159.
- Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. Memory and Cognition, 1, 19-40.
- Remington, F. J., & Joseph, A. J. (1961). Charging, convicting, and sentencing the multiple criminal offender. Wisconsin Law Review, 13, 528-565.
- Reyes, R. M., Thompson, W. C., & Bower, G. (1980). Judgmental biases resulting from differing availabilities of arguments. Journal of Personality and Social Psychology, 39, 2-12.
- Riskey, D. R. (1979). Verbal memory processes in impression formation. Journal of Experimental Psychology: Human Learning and Memory, 5, 271-281.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 10). New York: Academic Press.
- Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. Journal of Personality and Social Psychology, 32, 880-892.
- Schneider, D. J., Hastorf, A. H., & Ellsworth, P. C. (1979). Person perception (2nd ed.). Reading, MA: Addison-Wesley.
- Smith, E. R., & Miller, F. D. (1983). Mediation among attributional inferences and comprehension processes: Initial findings and a general method. Journal of Personality and Social Psychology, 44, 492-505.
- Special issue: Simulation research and the law. (1979). Law and Human Behavior, 3.
- Stasser, G., Kerr, N. L., & Bray, R. M. The social psychology of jury

- deliberations: Structure, process, and product. In N. L. Kerr & R. M. Bray (Eds.), The psychology of the courtroom. New York: Academic Press.
- Sulin, R. A., & Dooling, D. J. (1974). Intrusion of thematic ideas in retention of prose. Journal of Experimental Psychology, 103, 255-262.
- Tanford, S., & Penrod, S. (1982). Biases in trials involving defendants charged with multiple offenses. Journal of Applied Social Psychology, 12, 453-480.
- Taylor, S. E., & Crocker, J. (1980). Schematic bases of social information processing. In E. T. Higgins, C. P. Herman & M. P. Zanna (Eds.), Social cognition: The Ontario symposium. Hillsdale, NJ: Erlbaum.
- Taylor, S. E., & Fiske, S. T. (1975). Point of view and perceptions of causality. Journal of Personality and Social Psychology, 32, 439-445.
- Taylor, S. E., & Fiske, S. T. (1978). Salience, attention and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 11). New York: Academic Press.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.
- United States v. Foutz. Federal Reporter (2nd series), 1976, 540, 733-740.
- Vidmar, N. (1972). Effects of decision alternatives on the verdicts and social perceptions of simulated jurors. Journal of Personality and Social Psychology, 22, 211-218.
- Wyer, R. S. (1974). Changes in meanings and halo effects in personality impression formation. Journal of Personality and Social Psychology, 29, 829-835.
- Zadney, J., & Gerard, H. B. (1974). Attributed intentions and informational selecting. Journal of Experimental Social Psychology, 10, 34-52.

**END**