FORMAL AND EMPIRICAL RESEARCH ON CASCADED INFERENCE IN JURISPRUDENCE

DAVID A. SCHUM*
ANNE W. MARTIN

This paper reports observations from a series of formal and empirical studies of the process of assessing the probative value of evidence in the cascaded or hierarchical inference tasks commonly performed by fact finders in court trials. The formal research develops expressions that prescribe how the ingredients of various forms of evidence can be coherently combined in assessing the probative value of evidence. These expressions allow identification and systematic analysis of a wide assortment of subtle properties of evidence, many of which are commonly recognized in evidence law. The reported empirical research was designed to evaluate the consistency with which persons actually assess the probative value of evidence when they are asked to make these evaluations in several equivalent ways. Results show that persons, when required to mentally combine a large amount of probabilistic evidence, exhibit certain inconsistencies such as treating contradictory testimony as corroborative testimony and double-counting or overvaluing redundant testimony. However, when people are asked to make assessments about the fine-grained logical details of the same evidence, these inconsistencies do not occur.

I. INTRODUCTION

This paper contains an introduction to some of the results and observations we have accumulated from a series of studies on the task of assessing the probative value of inconclusive or probabilistic trial evidence. Some of these studies are formal or logical in nature and concern the manner in which the probative value of evidence *should be* assessed coherently. The other studies are empirical and behavioral in nature and concern the manner in which persons *actually do* assess the probative value of evidence. Our dual concern was voiced by Wigmore (1937: 8), who expressed interest in ". . . the reasons why a total mass of evidence does or should persuade us to a given conclusion, and why our conclusions would or should

---

* The research reported in this paper was supported by The National Science Foundation under Grants SOC 77-28471 and SES 80-24203 to Rice University. The authors gratefully acknowledge the advice and assistance during the planning of our research of Professor Richard Lempert, University of Michigan School of Law, and Professor L. Jonathan Cohen, The Queen's College, Oxford University. The authors also wish to thank Dr. Felice Levine, National Science Foundation, for her encouragement and support.

LAW & SOCIETY REVIEW, Volume 17, Number 1 (1982)

have been different or identical if some part of that total mass of evidence had been different." Our formal and empirical studies have proceeded hand-in-hand. Formal research helps to identify meaningful variables and measures for empirical research; empirical research, interesting in its own right, is also useful in testing the adequacy of the foundations for formal studies.

A major focus of our research has been upon inductive inference tasks, which Wigmore termed "catenated"; the modern terms for these tasks are "cascaded" or "hierarchical" (Wigmore, 1937: 13). Wigmore was the first to point out the fact that most inferential reasoning tasks are cascaded in nature. A cascaded inference task is composed of one or more reasoning stages interposed between evidence observable to the fact finder and the ultimate facts-in-issue. An example of a cascaded inference is presented by testimony from a witness of less than perfect credibility that the defendant was at the scene at the time of the crime. The testimony requires one first to assess the likelihood that the defendant was, in fact, at the scene/time. This foundational stage involves evaluation of the witness's credibility. Then, assuming the defendant at the scene/time of the crime, one must assess how strongly this event bears on the issue of whether or not the defendant committed the crime. Further difficulty is presented by intricate patterns of reasoning which require the joint consideration of current evidence with one or more previously given pieces of evidence.

The formal research discussed in this paper concerns the logical requisites of cascaded inference tasks and the manner in which these requisites should be combined. Wigmore acknowledged that logicians had found canons of reasoning in simple situations; however, he lamented the fact that (at that time) there were no such canons of reasoning from an entire mass of evidence "mixed" with respect to logical form (Wigmore, 1937: 8). Our formal work shows that, though the process is tedious and difficult, canons of reasoning can be derived for masses of mixed evidence.

The present article is meant to introduce the reader to ways in which formal analysis can be used to understand the logical demands of inference from the types of evidence presented in legal settings. As a result, the detailed mathematical arguments that form the bases for the analyses will not be described here. Interested readers are directed to our monographs dealing with these mathematical arguments.[1] Our focus in this article is on the major conclusions of our formal studies as they bear upon commonly encountered evidentiary issues in inferences made at trial. A technical appendix is included in which we illustrate the essentials of our formal process using three examples. The reader choosing to disregard this appendix is in no way disadvantaged in reading the text of this paper.

The benefits of formal reasearch on legal inference have not gone unnoticed by current scholars in jurisprudence (e.g., Lempert and Saltzburg, 1977; Lempert, 1977; Eggleston, 1978). Such research assists in efforts to illuminate and sharpen legal reasoning. The reader is, of course, interested in how our current research adds to this process.

Our empirical research concerns the reasoning processes of the ordinary citizens upon which so much depends in court trials. In fact, very little is known about human inferential reasoning, partly because of the lack of knowledge about the tasks people are asked to perform. Our formal research concerns what these tasks demand, and our empirical research concerns how well persons seem to meet these demands. Previous studies have suggested that everyone is subject to biases and errors in inferential behavior (e.g., Saks and Kidd, 1981). Unfortunately, many conclusions about human biases and error rest upon studies incorporating ill-posed problems or problems which are quite abstract. An objective in our empirical research was to present carefully posed concrete inferential problems which begin to approach the complexity of those faced by the fact finder in a court of law.

II. FORMAL RESEARCH: FOUNDATIONS

The inference tasks performed by the fact finder commonly involve the interplay of inductive and deductive reasoning processes; this fact was noted by Wigmore in his analysis (1937: 20). Our formal research generally focuses upon inductive inference, the task of revising one's opinion about the relative likelihood of rival facts-in-issue on the basis of inconclusive or probabilistic evidence. In evaluating or "weighing" evidence in an inferential task, one recognizes that items of evidence differ in strength; for various reasons some items are persuasive and allow for substantial revision in our opinions, while other items

[1] Requests for reprints or preprint monographs should be sent to the authors at: 7416 Timberock Road, Falls Church, Virginia 22043.

seem to justify little or no opinion revision. Thus, a major task in inductive inference consists of evaluating the inferential or probative strength of evidence. Given more than one item of evidence, one must somehow aggregate or combine the probative weights given to each item. One major complication is that the probative weight given to one item frequently depends strongly upon our recollection of other items. We explore the tasks of evaluating the probative strength of individual items of evidence and of collections of all the items. In essence, this research is concerned with relevance issues.

Federal Rule of Evidence FRE 401 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." As others have noted (Lempert and Saltzburg, 1977), there appear to be "natural" measures within conventional probability theory for the relevance or probativity of evidence. These measures are termed "likelihood ratios," and they provide an indication of the necessary opinion revision prescribed by FRE-401 for relevant evidence. In our context, a likelihood ratio expresses the probability of an item of evidence assuming the defendant's guilt, relative to the probability of this same item of evidence assuming defendant's innocence.[2] We use an upper-case Greek lambda [Λ] to symbolize a likelihood ratio and add a subscript when Λ applies to a certain evidence item e; thus, Λ_e means the likelihood ratio for evidence item e. When $\Lambda_e = 1.0$, evidence e has equal probability assuming guilt or assuming innocence, and so e is nonprobative. When Λ_e is greater than 1.0, then e is probative evidence favoring defendant's guilt; when Λ_e is less than 1.0, then e probatively favors defendant's innocence.

Our first formal objective has been to formulate and study likelihood ratio expressions for various identifiable logical species of evidence. We have termed these species of evidence *inference structures* and have used these inference structures as the basic "building blocks" in thinking about complex masses of evidence representing entire cases. The complexity of a likelihood ratio depends upon the form of the evidence and the nature of the reasoning process established by the evidence. Once derived for the evidence in some inference structure, likelihood ratio expressions tell us what probabilistic

---

2 Likelihoods and likelihood ratios appear as ingredients in Bayes' rule. Essentially, these ingredients prescribe the amount of revision, from prior opinion to posterior opinion, which an item of evidence justifies.

ingredients are required at each step in the reasoning process and how they should be combined in a coherent manner.

To many persons, equations seem sterile and, when applied as representations for human tasks, seem almost certain to fail in capturing all of the behavioral essentials that intuition and experience suggest are features of the tasks. Consequently, the second objective in our formal research has been the study of what we have termed the "behavioral richness" of our likelihood ratio expressions (Schum, 1977a). This term does not refer to the extent to which any of our formal expressions describes or predicts the actual behavior of any person. In our usage, a likelihood ratio is behaviorally rich if it captures the essentials and subtleties of probative value assessment that recorded experience with the evidence of concern suggest are there.

A third objective in our formal research has been to relate the study and analysis of likelihood ratio expressions for various inference structures to the rules and prescriptions of evidence law, noting both similarities and differences between the prescriptions of our formal expressions and corresponding established legal prescriptions. In general we have been impressed by the many similarities, a full accounting of which appears in our specific monographs. We view this correspondence between formal and legal prescriptions concerning relevance issues as evidence for a convergence to coherence in the development of evidentiary rules and procedures.[3] In addition, our formal research provides some basis for the sharpening of the definition of legal terms. As laypersons in jurisprudence, we have observed apparent difficulties among jurists in obtaining crisp definitions of certain terms (e.g., redundancy, corroborative and cumulative evidence). Formal research forces one to be precise, or at least to settle upon definitions.

[3] We note that our formal approach provides just one set of standards against which the coherence of legal prescriptions can be evaluated. There are other standards which we both recognize and appreciate. Our formal research is grounded on the axioms of "conventional" mathematical probability theory. There are other axiom systems which lead to other prescriptions for "coherent" inductive reasoning. One in particular is Cohen's system of "inductive" probabilities (Cohen, 1977). Cohen claims this system to be more congenial to application in legal matters than the mathematical systems which we use; much current debate on this issue has resulted (e.g., Schum, 1979a; Kaye, 1979, 1981; Cohen, 1981). Other systems include the "belief functions" of Shafer (1976), and the "possibility" measures of Zadeh (1978). We continue to work within the mathematical system, because it is our belief that this system is the only one extant which offers the flexibility necessary to capture the rich array of subtleties in evidence.

One of our strong hopes has been that our formal research will assist us in efforts to perform empirical research of greater interest and relevance to jurists. Laboratory research on human inference is often criticized for being too abstract and for not including enough relevant aspects of tasks as they occur in natural settings (Winkler and Murphy, 1973). In our empirical research, subjects evaluated evidence having a variety of subtle properties. Likelihood ratios, derived for the evidence our subjects evaluated, contain ingredients which, far from being mathematically arcane, in fact lay bare the logical steps or stages in reasoning from the evidence to the major facts-in-issue.

III. FORMAL STUDIES: RESULTS AND OBSERVATIONS

On Inference Structures

Substantively, evidence varies in near-infinite fashion. Fortunately, however, there appears to be a manageably finite number of logically distinct forms of evidence. Various classification schemes are found in evidence law treatises, but no one scheme seems to enjoy universal acceptance. Our formal work has basically involved circumstantial evidence with a testimonial foundation; Wigmore noted that such evidence is the form most frequently encountered at trial (1937: 13). Our formal process works equally well for "real" evidence; in fact, formalizations for the probative value of such evidence are simpler than for testimonial evidence (Schum, 1980). Our interest in testimonial evidence arises partly because of an abiding interest in the nature of the relationship between the credibility of the source of evidence and the inferential or probative value of what the source reports (Schum and Du Charme, 1971).

There appears to be a relatively small number of generic types of evidence which we have termed *inference structures*. We distinguish between simple, complex, and mixed inference structures. A *simple* inference structure is a chain-like reasoning process "set in motion" by a single item of testimony. The foundation of the reasoning chain is a testimonial assertion; later stages or links in the chain represent circumstantial reasoning steps from the matter asserted to the ultimate or major facts-in-issue. In such structures the number of reasoning steps can vary. A further characteristic of a simple inference structure is that neither the testimonial assertion nor the events at subsequent reasoning stages are assumed to be probabilistically linked to previous evidence or events in their reasoning chains. Shown in Figures 1-A and 1-B below are diagrammatic representations of two simple inference structures involving testimonial evidence. The one in Figure 1-A, termed a "first-order cascaded inference," is the simplest possible case of cascaded or catenated inference. Witness W_i asserts that event D happened—for example, that the defendant was at the scene of the crime. The event D^c is the event that D did not occur. We represent the testimonial assertion that D occurred as D_i^* to distinguish it from the event D itself. Failure to distinguish between testimony about an event (D_i^*) and the event itself (D) has caused no end of difficulties in many previous studies of the impact of witness credibility upon the probative value of testimonial evidence (e.g., Eggleston, 1978). The first stage of reasoning is from testimony D_i^* to events D, D^c; this is the foundation stage of reasoning and, as formalizations for the probative value of D_i^* show, involves assessment of the credibility of W_i. The next stage of reasoning is from events D and D^c to events H_1 and H_2, representing the major or ultimate facts-in-issue (e.g., that defendant committed the crime, H_1, or did not, H_2). In this stage of reasoning the probative

Figure 1. Example Inference Structures

A. Simple cascaded inference, first order.
B. Simple cascaded inference, second order.
C. Complex cascaded inference, corroborative testimony.

A.

(H_1, H_2)
↑
(D, D^c)
↑
D_i^*

B.

(H_1, H_2)
↑
(D, D^c)
↑
(E, E^c)
↑
E_j^*

C.

(H_1, H_2)
↑
(D, D^c)
↑ ↑
D_i^* D_j^*

https://doi.org/10.2307/3053534 Published online by Cambridge University Press

importance of the defendant's being at the scene of the crime is assessed. Equation 1 in the technical appendix shows the expression for the likelihood ratio for testimony D_i^*.

Figure 1-B shows a "second-order cascaded inference." Witness W_j asserts E_j^*, that event E occurred. Suppose E is the event that the defendant's car was at the scene of the crime at the time in question. The first stage of reasoning, from E_j^* to events, E,E^c, is an assessment of the credibility of W_j. The next stage involves circumstantial reasoning from E,E^c to events D,D^c whether or not defendant was at the scene. The final stage involves circumstantial reasoning from events D,D^c to events H_1, H_2. Equation 2 in the technical appendix shows the likelihood ratio for testimony E_j^*. Simple inference structures like these two can have any number of reasoning stages or "levels" of cascading or catenation.

In *complex* inference structures there is always foundation testimony from more than one witness. Often there are probabilistic linkages among events in the reasoning chains based on each item of testimony. We have identified four basic classes of complex inference structures: those representing *contradictory* testimony, *corroboratively redundant* testimony, *cumulatively redundant* testimony, and *nonredundant* testimony. In contradictory testimony, one witness asserts that event D occurred and another asserts that D did not occur. Corroboratively redundant testimony concerns the assertions of two or more witnesses that the (same) event D occurred. Figure 1-C depicts a complex inference structure involving corroboratively redundant testimony. In this structure, two witnesses W_i and W_j assert that event D occurred; their testimonies are D_i^* and D_j^*. That the testimony here is possibly redundant is obvious, since they both testify to the same event. If the first witness has perfect credibility, then testimony from the second usually adds nothing. Suppose, however, we thought that the first witness could not actually determine whether or not D occurred. Then, testimony from the second witness does have probative value depending, in part, on the credibility of this second witness. This situation introduces the important feature of *conditional nonindependence*. Two or more items of evidence or events at reasoning stages suggested by evidence are conditionally *non*independant if, considered jointly, they mean something probatively different than they do if considered separately. If not, then they are said to be conditionally independent.[4] Equation 3 in the appendix shows the likelihood ratio for the second and possibly redundant testimony. There are terms in this expression which allow one to account, in a formally ideal way, for the degree of redundance involving these two witnesses.

In *cumulatively* redundant testimony, there is an assertion from one witness that event E occurred. A later assertion that a different event F occurred comes from either the same or a different witness. Suppose that event E, if it occurred, made event F highly probable in the nature of things (i.e. regardless of what else you know), and therefore testimony that F occurred yields little if any probative value if the first witness is believed. For example, the first witness asserts that defendant was at the scene/time of the crime. The second asserts that he/she found the defendant's coat at the scene/time. If the first witness is believed, testimony from the second adds little to our determination about whether or not the defendant was at the scene/time. The redundance of cumulative testimony of this sort can be represented, along with other subtleties, in an appropriate expression of conditional nonindependence.

If two or more items of testimony are redundant, earlier items tend to decrease the probative value of later items. Alternatively, testimony can be *facilitative* so that one item makes a later item seem more probatively valuable. In *nonredundant* testimony, either the assertion of one witness causes no change in the value of a later assertion, or it acts to enhance the value of a later assertion. For example, the first of

---

[4] At various points we use the expression conditional independence or nonindependence of events, because with it we can represent a remarkable array of subtleties in evidence. Two events A and B are said to be unconditionally independent or simply *independent* if knowledge of one of the events does not cause you to change your mind about the probability of the other; if such knowledge does cause a change, then the events are said to be nonindependent. Suppose we have knowledge of a third event C. Events A and B are said to be *conditionally independent*, given event C, if knowledge of one of the events A or B does not cause you to change your mind about the probability of the other, provided that event C is true. If knowledge of event A causes you to change your mind about the probability of B, when event C is known or assumed, then A and B are said to be *conditionally nonindependent*, given event C. Very often, two or more evidence items or events at reasoning stages suggested by these items mean something probatively different when considered jointly than they do if considered separately. The concept of conditional independence/nonindependence allows you to express this probative difference. Unfortunately, the concepts of independence and conditional independence are often confused; they are related concepts, but they are not the same. Examples of subtleties which find expression via patterns of conditional nonindependence include a variety of credibility-related effects, redundance in testimony, the significance of weak links and rare events in reasoning stages, a reasoning stage relation called transitivity, and the locus of probative value in equivocal testimony or the nonproduction of testimony.

two witnesses reports finding the defendant's revolver beside the victim's body in the defendant's apartment. The second witness reports the sound of a revolver shot coming (apparently) from the window of this apartment. On the issue of whether or not the defendant is guilty, the first testimony seems to have a facilitative effect upon the probativity of the second.

Finally, a *mixed* inference structure represents various combinations of the above structures. As an example, two witnesses testify that event E occurred; their testimony is corroboratively redundant. Then, three witnesses give joint corroborative testimony that event E did not occur, thus contradicting the testimony of the first two. The reader who is interested in a complete collection of the inference structures we have studied, including a derivation of likelihood ratio expressions for each structure, may refer to Schum and Martin (1980a; 1981).

Cases or collections of evidence can be thought of as collections of inference structures. Testimony or other evidence upon which these structures are based open up specific lines of reasoning which the structures indicate. Such evidence may be thought of as "main-frame" evidence, having direct probative value. Other evidence may be thought of as "ancillary" evidence in the sense that it allows the fact-finder to evaluate the strength of linkages among events at various stages in the reasoning suggested by an inference structure. Ancillary evidence may be said to have "acquired" or "derived" probative value. For example, Witness W_j asserts that the observational conditions were good on the day that a previous witness W_i observed E, an event linked circumstantially to major facts-in-issue. The testimony of W_j is not probative on these facts by itself; it does, however, acquire probative value since it bears on the credibility of W_i, whose testimony about E does have direct probative significance. Thus, all evidence at trial can be grossly categorized as "main-frame" or "ancillary" evidence. Had Wigmore realized this, he might have been able to simplify some of his very complex diagrammatic illustrations of case evidence, and he might also have been able to see how specific formalization of such evidence could be derived.

On Witness Credibility

The credibility of the source or sources of the evidence forms the foundation for cascaded reasoning from testimonial evidence. Likelihood ratio expressions for the process of

https://doi.org/10.2307/3053534 Published online by Cambridge University Press

assessing the probative value of testimonial evidence reveal several important logical characteristics of this process. First, established grounds for impeaching or supporting the credibility of witnesses find expression in these likelihood ratios (Schum and Kelley, 1973; Schum, 1977a). These grounds include observational capacity, bias, prior inconsistent statements, influence among witnesses, and character (Cleary, *et al.*, 1972). Second, our formal process makes clear the fact that the behavior of a witness, as revealed by the witness or by other evidence, is often a source of probative value in addition to that provided by the events reported by the witness. Finally, our studies show the precise nature of the important interaction between the credibility of a witness and the rarity of the event reported by the witness in determinations of the probative value of witness testimony (Schum, 1977a).

In studying the probative value of witness testimony we have found it useful to use two constructs from signal detection theory, a theory which has had great impact on sensory psychophysics and a variety of other research areas (e.g., see Swets, 1964; Egan, 1975). This theory provides the means for separating two basic dimensions of testimonial behavior. The first dimension, which concerns the observational sensitivity or capacity of a witness, is indexed by a measure labeled d′. The second dimension concerns motivational and other factors that influence the decisions by a witness about what event to report following an observation. This decision criterion, $L(x_o)$, can be determined from information about the observer's expectations, goals, and motives.

The hit and false-positive probabilities in our Λ formulations, together with other directly related probabilities called "misses" and "correct rejections" are key elements in signal detection theory; in fact, these labels come from this theory. For some observational task, if we know or assume the conditional probability of an event being reported not true (a false positive), we can determine d′ and $L(x_o)$. Moreover, we can vary d′-related information independently of $L(x_o)$-related information in the formal study of how such classes of information influence the probative value of testimonial evidence.

In our formal studies we first examined situations in which credibility-related hit and false-positive probabilities were not conditioned by other events in a reasoning chain. When this is true, straightforward trade-offs are possible between observational sensitivity and decision-related factors in determining the probative value of testimony. For example, testimony from a witness with low observational sensitivity but a strong bias against offering the testimony can have as much value as testimony from a highly sensitive witness who is biased in favor of giving such testimony. We also attempted to clarify the meaning of "biased" testimony in relation to testimony that lacks veracity. In our context, bias refers to a witness' apparent preferences for or against offering the testimony. Such preferences can sometimes be inferred from other evidence such as information about the relationship between witness and defendant. Suppose a witness is a close friend of defendant; this witness may have a distinct bias in favor of reporting an event favorable to the defendant's case. We may easily believe the witness to be biased without also believing that the witness is lying when he/she reports the occurrence of this event. Our formal process hows why biased testimony need not be untruthful, and how testimony which lacks veracity need not be biased. Generally, bias is a factor in the determination of the probative *strength* of testimony, while veracity determines the probative *direction* of testimony (i.e., which of the two rival facts-in-issue the testimony favors).

We also examined a variety of situations in which observational capacity and/or decision-related factors were made conditional upon events representing rival facts-in-issue. This is another way of saying that credibility-related factors for a witness provide probative value over and above the value of the event being reported. As an example of how factors underlying the observational and decision-related factors influence the value of testimonial evidence, we have examined a case in which a witness testifies against preference. In such a case one expects a "gain" in probative value over identical testimony from a witness who has no such preferential bias. Our formal studies show that this gain is jusitified and show the precise formal locus of such gain. In general, the observational and testimonial behaviors of a witness by themselves are important sources of probative value. These studies show just how important these sources are, since our formulations are remarkably sensitive to apparently minor alterations of credibility-related ingredients when they are made conditional upon one or the other facts-in-issue.

Redundance Issues

Redundance is among the most interesting but formally difficult evidentiary issues. Our interest in the formal study of the redundance of testimonial evidence was stimulated by Lempert's (1977) concern about the "double-counting" of such evidence. In our studies (Schum, 1979b) we distinguished between "cumulative" and "corroborative" redundance since we observed a formally necessary distinction between the instances in which two or more witnesses say the same thing and the instance in which two or more witnesses give different testimony but on obviously related matters. Our usage of the term *redundance* corresponds with the common interpretation that redundant evidence is superfluous and supplies little, if any, probative value in addition to previous evidence. We explored the various uses of the terms "corroborative" and "cumulative" and observed some confusion; some jurists make sharp distinctions between these terms, while others use them interchangeably. In our study we found it necessary to make a distinction since, formally, corroborative redundance is a special case of cumulative redundance. In our studies of the factors which influence the redundance of testimonial evidence, it is apparent that there are more factors influencing cumulative redundance than there are influencing corroborative redundancy.

In a strict sense *redundance* is a property of the *events* being reported and *not* the testimony of these events. To see that this must be so, consider the testimony of two witnesses who both report that event D occurred. The first witness, we are convinced, could not tell whether or not D occurred, and therefore we assign *no* probative value to this testimony. The second witness seems reasonably credible, and we are justified in assigning probativity to this second testimony to the extent that the second witness is credible and to the extent that event D is probatively valuable.[5] Equations 4 and 5 in the technical appendix express our measures of event redundance in the cumulative and corroborative cases.

The following six factors influence the *cumulative* redundance of testimony F_j^*, given prior testimony E_i^*: The strength of the redundance of *events* E and F (as measured by R_{cum} in Equation 4 in the technical appendix), the credibility of W_i and of W_j; the probative value of event F, given E^c (the first witness may be untruthful); the rarity of events E and F; and

[5] Concepts from *statistical communications theory*, often called *information theory*, allow us to measure the probative redundancy of events in well-defined reasoning chains. In this theory, measures of redundancy are crucial in assessing the efficiency of ideal and actual communications channels (e.g., Staniland, 1966). Our measure of *event* redundance has essentially the same properties as does the redundance measure in information theory.

the number of reasoning stages between E and F and major facts-in-issue. Four factors determine the redundance of *corroborative* tetimony E_i^* and E_j^*: the credibility of W_i and of W_j; the rarity of event E; and the number of reasoning stages between E and the major facts-in-issue. Careful study reveals a large number of interesting consequences to probativity assessment when redundancy is ignored. For example, ignoring redundancy of E and F will sometimes, but not always lead to the overvaluation of testimony F_j^*. Certain combinations of credibility-related ingredients for witnesses W_i and W_j cause one to *undervalue* testimony F_j^* even when events E and F are strongly redundant. In the corroborative case, ignoring the natural redundance of such testimony is most serious when the credibility of the prior witness W_i is strong; i.e. "double-counting" of evidence in the corrobative case is most serious because, when the prior witness is highly credible, the value of testimony from the second witness is, formally, nearly valueless. The rarity of events exerts an interesting influence in the corroborative case. When the event being reported is rare, a stronger level of credibility of the first witness is required to make the value of the testimony of the second witness vanish when natural corroborative redundance is ignored.

Reasoning Chain Issues

In the analysis of inductive reasoning chains, a variety of issues arouse considerable interest. Following are three issues which we have examined using various analyses of likelihood ratio formulations for simple inference structures based upon a single item of foundation testimony (Schum, 1979c). The first issue concerns a formal relation called *transitivity*. Suppose foundation evidence A probatively favors B, and B probatively favors C; this chain of reasoning is said to be *transitive* if A probatively favors C. If A does not favor C, the relation in the chain of reasoning is *intransitive*. We have examined a variety of conditions under which such transitive relations are either formally allowable or denied. The second issue concerns the effects upon the probative value of foundation testimony of locating a "weak link" at various points in an inferential reasoning chain. The third issue concerns the effects upon the probative value of foundation testimony of locating a "rare" event at various points in a chain of inferential reasoning. Transitivity and weak link issues and their analysis within the conventional probability system were critically examined by

Cohen (1977). Our study of these issues was prompted, in part, by Cohen's analysis, since we did not believe that all problems raised were adequately posed.

Study of conditions favoring transitive relations among stages of inductive reasoning are particularly important in civil cases in which at least some courts enforce different proof standards at different stages of an inferential chain (Cohen, 1977: 69). For example, a foundation stage may require a more stringent standard of proof in order to support a subsequent stage whose proof standard is "on balance of probabilities" or "preponderance of the evidence." Our analysis shows that whether or not transitive relations occur formally in a chain of reasoning depends entirely upon the pattern of conditional independence/nonindependence relations among events in a reasoning stage. Suppose there is a complete pattern of conditional independence relations among events in a reasoning stage; this means that any event is conditioned only by events at the next higher stage of reasoning. When this is true, if foundation A probatively favors B, and B probatively favors C, then A will probatively favor C. However, under various patterns of conditional nonindependence in which an event at one stage may be linked with those at several higher stages, transitive relations, though intuitively expected, are frequently denied by appropriate formalization. Thus, it is not true in general that, if argument at each reasoning stage favors side A, then the overall argument favors A.

As an example, suppose testimony that the defendant was at the scene (D_i^*) favors the event that the defendant actually was at the scene (D) and, in turn, the defendant being at the scene (D) favors the event that the defendant is guilty as charged (H_1). Our formal process shows why it is not necessarily the case that testimony that defendant was at the scene (D_i^*) probatively favors defendant's guilt (H_1). This probative relation may be transitive or intransitive, depending on what we know about the witness.

Analysis of "weak links," "rare events," and their location in a reasoning chain also reveals the importance of conditional independence/nonindependence patterns among events in a reasoning chain. Some may expect that a weak link at the foundation stage of an argument is more damaging than one located at the "top" of a chain. Grounds for this expectation seem to be that a strong foundation for a weak argument is preferable to a weak foundation for a strong argument. Our formal analysis shows that neither is preferable provided that there is a pattern of complete conditional independence among events in a reasoning chain. This says that the location of a weak link does not matter under these conditions and that the chain cannot be stronger than its weakest link wherever it is located. The location of "rare" events in a chain *does* matter. Rare events are more damaging to the probative value of foundation testimony if they occur at the top of the chain, whether or not there are conditional nonindependence patterns in the chain.

On Several Special Types of Evidence

Following is a brief account of formal issues encountered in our examination of three well-known types of evidence. Study of these types of evidence, all of which can involve cascaded inference, has required us to examine several issues of more general importance in the analysis of complex reasoning chains.

Hearsay Evidence One of the most difficult tasks in the formal study of evidence concerns a type of evidence that each one of us evaluates on a regular basis; in jurisprudence such evidence is termed "hearsay"; more generally it is called "second-hand" evidence. This latter designation is more general than it appears, since it is usually applied to instances in which one receives a report or testimony passed through several intermediate sources; often, the original source is unknown. In evidence law there is an abundant literature on the admissibility of various forms of hearsay but a sparse literature on the probative value of such evidence. Suppose a "simple" situation in which A, a witness at trial, reports testimony allegedly given by out-of-court assertor B. There are, of course, complex credibility-related issues concerning both A and B; so, at the very least, we have more than one foundation stage in reasoning from such evidence.

Our formal studies of hearsay (Martin, 1979) represent elaboration and extension of Tribe's model for the "triangulation" of hearsay in which only admissibility issues were of concern (Tribe, 1974). Likelihood ratio representations for the probative value of hearsay are made difficult by the fact that there are several possible different stages of reasoning involved. We may ask: did B actually observe the event in question; did B in fact report anything to A? (A may have "put words into B's mouth".) Derivation of Λ for alternative forms of reasoning from hearsay evidence led to the discovery of some very important *recursive* algorithms which occur in all Λ formulations. A recursive rule is one that is defined in terms of itself. Such discovery led to the development of a *general* algorithm for the analysis of inference structures of virtually any degree of complexity and involving any pattern of conditional independence assumptions involving events in the structure (Martin, 1980). A computer program called CASPRO has been developed which uses this algorithm and which facilitates analysis of complex inference structures.

Equivocal Testimony Or No Testimony On A Relevant Issue at Trial We now consider three species of "evidence" which some, at least, would judge to be, by their very nature, probatively vacuous regardless of the facts-in-issue (Schum, 1981b). Formal study, however, convinces us that there is the possibility of very strong probative value in each case. The first concerns *equivocal testimony* given by a witness who, when asked whether or not the event D occurred, replies "I don't know," "I don't remember," "I couldn't tell," etc. In some instances this may simply indicate a form of self-impeachment by the witness who, in fact, is truthful in giving any of these responses. Formally, such testimony is probatively valueless. In other instances, depending upon what else we may know about the witness, we may infer that the witness is "sandbagging" and actually knows more than his/her testimony indicates. In such instances, formal analysis shows that the probative value of such equivocation can be even more probative than *certain* knowledge of the occurrence of one of the events about which the witness equivocates.

The second case concerns *silence* as evidence; queried about whether or not event D occurred, the witness stands mute or exercises privilege. In this case other facts brought to light at trial can, under some conditions, formally justify stronger opinion revision about facts-in-issue than would specific testimony by the witness. The same also applies to a third case involving the *nonproduction* of evidence (testimonial or otherwise). Such nonproduction can be at least as probative as specific evidence on the matter at issue given by a witness of any level of credibility.

Opportunity And Alibi Testimony The array of subtleties in evidence make the formal study of evidence rather like walking through a mine-field; one wrong step, however minor, can cause subsequent discomfort. We experienced such discomfort in applying our formal processes to opportunity evidence and its logical negation, alibi evidence. Probatively, opportunity evidence, even if given by a perfectly credible source, is inconclusive that the defendant committed the act as alleged. Alibi evidence, if given by a perfectly credible source, is conclusive evidence of defendant's nonparticipation in the act as alleged. We assume here that the act in question would, if performed by defendant, require his/her physical presence. Unless Λ equations are formulated with extreme care, certain embarrassing indeterminacies can arise in equations whose ingredients seem entirely plausible unless carefully examined (Schum, 1981c).

As a final word on formal issues we mention Wigmore's desire for "mental probative equations" (Wigmore, 1937: 8). Whether or not our Λ formulations bear any resemblance to the equations Wigmore had in mind, we can never tell. Our formal study reveals the intricacies of cascaded or catenated inference even for simple items of evidence. The process of "connecting up" the evidence is, formally, frightening to contemplate. We have done so, however, for small evidence sets. This process, we have noted, bears no small resemblance to the sensory-perceptual tasks of observing an object against a background. Prior evidence exerts a "contrast" effect on current evidence in the same way as the color of a background influences the perceived color of an object presented against this background (Schum, 1977b). Before considering empirical research issues we note our realization that formal study of evidence by itself can never prescribe what the "rules of evidence" ought to be. However, as one jurist (Keyser, 1929) remarked:

> An ensemble of experience-given propositions (like those constituting any existing branch of law) never gets so thoroughly examined and criticized and understood as when the ensemble is submitted to the severe processes of mathematicization.

IV. EMPIRICAL STUDIES: BACKGROUND, OBJECTIVES, AND METHODS

Our empirical research was designed to provide information about human capabilities and limitations in the task of assessing the probative or inferential value of evidence. In a series of related studies whose results are summarized below, research subjects provided specific numerical responses as indications of their assessments of the probative strength of individual items of evidence and of collections of evidence items. Examination of these responses in various ways is one means of studying certain characteristics of human behavior in the task of reasoning from inconclusive evidence. In actual court proceedings, however, fact finders are not encouraged to make public their judgments about the evidence. At the outset we note that no part of our study was designed to convince jurists that the fact finding process ought to require specific quantitative assessments of evidence strength or that forensic standards of proof ought to be specifically quantified. This disclaimer seems necessary in light of controversy among jurists about the use of various probabilistic representations in actual court trials (e.g., Tribe, 1971; Finkelstein and Fairley, 1970).

It is commonly believed that the quantitative reasoning skills of ordinary persons are not very strong and that quantitatively expressed human judgments are neither accurate nor reliable. In a variety of research contexts, psychologists and others have obtained useful results by relying heavily upon numerical judgments expressed by their research subjects; we do so here.[6]

Background and Specific Objectives

Our present studies of human capabilities and limitations in the task of assessing the probative weight of evidence have three major roots: Wigmore's work on the analysis of complex masses of mixed evidence, basic and applied research performed by psychologists and others concerning the design of more efficient ways to allocate tasks among persons and devices in various information-processing systems, and our recent work on the formal analysis of cascaded inference. We have already acknowledged our debt to Wigmore, who began the formidable task of decomposing complex inferences. Within psychology, study and analysis of complex inference tasks dates from the early 1960's with Ward Edwards' suggestions about how such inference tasks ought to be

---

[6] In the study of sensory and perceptual processes, for example, there is a large collection of *psychophysical* measurement procedures, many of which require numerical judgments from subjects. In sensory psychophysics, it is common practice to take seriously the numerical judgments individuals provide as indications of various attributes of their sensory experience. Such faith has not been unrewarded; a variety of useful metrics and measurement procedures based upon subjective quantitative assessments are employed on a regular basis in very practical applications related to vision and audition. Behaviorally useful measures of light and sound are all based upon psychophysical measures (e.g., Stevens, 1975). Our present studies are within the tradition of psychophysics in the sense that we take seriously the numerical judgments about evidence strength provided by our subjects. We do, in fact, construe these judgments as psychophysical judgments and, in so doing, are consistent with the philosopher Hume's assertion that all probabilistic reasoning is a species of sensation; the weight given to alternative arguments involves subjective feelings about the relative intensity or strength of the arguments (Hume, 1881).

decomposed so that a person, confronted with a mass of evidence, could be relieved of the task of mentally aggregating large amounts of probabilistic evidence (Edwards, 1962). As Edwards discussed, Bayes' rule is suggestive of ways in which such task decomposition ought to be performed. These ideas generated a substantial amount of research on matters concerning human performance on inductive or probabilistic reasoning. Most of this research was performed in laboratory settings involving tasks of varying complexity; excellent reviews are to be found in papers by Slovic and Lichtenstein (1971) and Rapoport and Wallsten (1972). Research on inferential processes continues; recent research on various attributes of human probabilistic judgments of concern to jurists has been reviewed by Saks and Kidd (1981). Our own formal research on cascaded or hierarchical inference was designed to extend the applicability of Bayes' rule to complex forms of evidence.

In planning our studies of cascaded inference in jurisprudence we had a number of objectives. Three of these objectives concern our basic interest in the task of assessing the probative weight of evidence, and the results presented in the next section bear upon these.

(1) In the study of many human tasks a natural question is: how accurately or correctly can a person perform the task? Unfortunately, as we recognize, we can *never* ask how correct or accurate is a person's assessment of the probative weight either of an item of evidence or a collection of evidence given at trial. Such evidence involves unique or one-of-a-kind events, and each fact finder evaluates the evidence according to personal strategies based upon a unique matrix of prior experience. In short, there is no "true" or "correct" probative weight for any item or collection of evidence; still, even though we cannot measure the accuracy of probativity assessment, we can, under certain circumstances evaluate the consistency or coherence of such assessments. One such circumstance occurs when a person can be asked to perform the same task in different but equivalent ways. As we shall see, a probativity assessment task can be decomposed to varying degrees; we can ask how consistent or coherent are assessments across various levels of task decomposition. In a word, this objective concerns the extent to which a person's assessment of the value of "parts" of an evidence collection are consistent with this person's assessment of the "whole" collection of the evidence. Our formal process supplies the essential basis for showing

how probative value assessment tasks can be decomposed in formally equivalent ways; this process also, of course, shows us how the decomposed "parts" ought to be put back together again.

(2) The evidence items evaluated by our research subjects are classifiable into the logically distinct categories that we have termed "inference structures." Our second objective was to study characteristics of responses to these forms of evidence. One might expect that the consistency with which persons evaluate evidence would depend upon the logical form of the evidence.

(3) Study of the internal consistency of probativity assessment involves a focus on the performance of individual persons. We were concerned, however, about the extent of the agreement or concordance across subjects in probativity assessment for various forms and collections of evidence.[7]

Methodological Choices And Trade-Offs

At a fairly early stage in our empirical research planning, we abandoned hope of being able to present precise descriptions of our methodology in any journal-length account of our work. Our solution to this problem was to prepare a monograph in which we provided a detailed account of all aspects of our empirical studies, including precise descriptions of the subjects' tasks, the exact instructions that they were given in each part of the study, the evidence they evaluated, the formal basis for selecting the evidence, the nature of the subjects' responses to the evidence, and the means by which these responses were to be analyzed. The report is available to persons interested in these details (Schum and Martin, 1980a). The reader of this report may conclude that we over-reacted to criticism about the simplicity of "laboratory" studies of human inference (e.g., Winkler and Murphy, 1973), and that we sought

[7] Other objectives also guided some of our research. One objective concerned a test of the relative adequacy of Pascalian (mathematical) and Baconian (inductive) systems in charting the general course of probability revision based upon inconclusive trial evidence. A novel methodology is required for such a comparison, because the two competing systems have almost no comparable numerical properties. The reader interested in our approach to and results of such a comparison can refer to our recent preliminary report (Schum and Martin, 1980b).

Another objective of the research concerned how subjects' perceptions of the value of trial decision consequences for a defendant may influence their assessment of the probative value of evidence. Our analysis of results bearing upon this objective, though at this time incomplete, suggests that value influences are very slight. This should be good news to decision and inference theorists in general, since most decision theories assume an absence of interaction among inference-related and value-related judgments.

complexity for its own sake, but this assessment would be inaccurate. We sought to study, in a reasonably systematic way, human inferential reasoning processes in an evidentiary context which begins, at least, to approach the evidentiary complexity of actual court trials. Quite simply, there is a price to be paid for the incorporation in research on human inferential reasoning of the subtleties in evidence upon whose recognition and evaluation in actual trials so much depends.

Following is a brief account of the necessary trade-offs we were forced to consider in the design of our empirical research. If one wishes to study a person's evaluation of collections of formally identifiable classes of evidence, one must either find actual cases whose evidence fits into these classes, or one must contrive evidence which does fit in these classes. After an unsuccessful attempt at the former, we chose the latter. Our study of the consistency with which individuals evaluate evidence requires that a person evaluate the same evidence over again on several occasions; possible "carry-over" effects from one evaluation to another, though they cannot be eliminated, can be minimized. We were forced to rule out two ingredients which may have increased at least the "surface validity" of our studies. The first concerns the use of a "random" sample of juror-eligible persons. Our study demanded lengthy subject participation over a period of 5½ weeks, ability of subjects to understand fairly detailed instructions, and absolute subject commitment to participate in all parts of the study. We considered the use of actors in a video presentation of "trial testimony"; this was rejected for reasons of time and expense. Consequently, our subjects received written accounts of "testimony." Since our major purpose was to examine individual evaluations of evidence, our studies did not involve a group deliberation process similar to actual jury deliberation. Finally, to make certain evidentiary events occur "on cue," we were forced to take some liberties with the natural order of evidence presentation as it might occur at trial. Following is a brief account of the details of our method.

Research Subjects Twenty jury-eligible persons, ten males and ten females, from the undergraduate population at Rice University completed all phases of our study. None of these persons had taken a college-level course in probability theory (the only stated requirement), and all were paid at an hourly rate with a bonus for completion of all phases of the study. At the completion of our studies, we had compiled a data base of over 16,000 numerical assessments provided by the twenty subjects in response to the evidence they were asked to evaluate.

The Evidence Our subjects evaluated *testimonial* evidence in 12 separate *contrived* felony "cases." Evidence from each case was presented in written "transcripts," each of which consisted of a description of the defendant and the crime with which he/she was charged, followed by several separate blocks of evidence. Evidence presented in each case was not intended to represent a complete case; this is not a crucial issue, since our research subjects were never asked to make judgments about the guilt or innocence of the "defendant" in any case. In fact, their only task was to assess, in various ways, how strongly case evidence favored the guilt or the innocence of the "defendant" in each case.

Each block of evidence in a case consisted of two basic forms of information. The first, which we have called "main-frame" evidence, is a testimonial assertion by a witness whose assertion opens up a line of reasoning to major facts-in-issue. The second, which we referred to as "ancillary evidence," concerns other evidence which bears upon the task of evaluating the strength of linkages in the reasoning chains suggested by "main-frame" testimony. Such ancillary evidence included evidence regarding the credibility of the witness and other explanatory evidence such as that brought out in cross-examination or by other rebuttal witnesses. In Wigmore's terms, each block of evidence consisted of *proponent's assertion* followed by *opponent's explanation and/or denial*. Each of the twelve "cases" contained either four, five, or six blocks of evidence.

The most important characteristic of each block of evidence was that it was contrived to fit into one of 15 well-defined inference structures such as the ones shown above in Figure 1. Thus, the essential logic of each "case" was represented as a collection of inference structures; the substance of the evidence in a case was contrived to fit the logic of each case. These twelve cases, together with their "logic diagrams" and likelihood-ratio equations appropriate in each inference structure in the logic diagram, are also available upon request to the authors. Our formal research has provided the means for developing likelihood ratio expressions (Λ) for the evidence in each inference structure used. There were five basic classes of inference structures, those representing: "simple" inference structures, contradictory testimony, corroboratively redundant testimony, cumulatively redundant testimony, and nonredundant testimony. The substance of the "main-frame" testimony concerned either means, motive, or opportunity evidence relevant to each case. We drew heavily upon Wigmore's examples in contriving the evidence in each structure (Wigmore, 1937), especially his many examples of the number of reasoning stages typically necessary for means, motive, and opportunity evidence. Typically, for instance, motive evidence requires more reasoning stages than does opportunity evidence, since a motive is inferred from past behavior. Finally, the likelihood ratios developed for each inference structure show the necessary probabilistic ingredients and the manner of their coherent combination in the task of assessing the probative value of evidence in each inference structure. As we now discuss, such formalizations, together with other elements of Bayes' rule suggest three equivalent formal means for assessing the probative value of the evidence in each case.

Research Subjects' Tasks And Responses Subjects provided assessments of the probative value of evidence in three formally equivalent ways; these three response methods allow us three ways of assessing the internal consistency of evaluations of evidence for entire cases or collections of evidence and one way of assessing the internal consistency of evaluations of individual items of evidence. Two of these three response methods involve the estimation by subjects of likelihood ratios; the third involves the estimation of conditional probabilities. Following are the three response methods for probative value assessment used by every subject for each "case."

Zero Task-Decomposition (ZTD) In this response method, subjects estimated a single likelihood ratio for the *entire* collection of evidence in a case. In giving such an estimate, a subject is asked to assess the likelihood of case evidence assuming defendant's guilt relative to the likelihood of this same evidence assuming innocence. Subjects' actual responses consisted of a letter-number pair which indicated both the probative *direction*, i.e., whether the case evidence favored guilt G or innocence I, and probative *force*. For example, the pair G-10 indicates that the subject thinks the evidence 10 times more probable assuming guilt than assuming innocence;

the pair I-5 indicates that the subject believed the evidence 5 times more probable assuming innocence than assuming guilt. A response "N" meaning "neutral" was allowed if the subject believed the evidence was probatively neutral and favored neither fact-in-issue G nor I. Our computer converted these pairs to ratios according to the following definition of the likelihood ratio for case evidence C: $\Lambda_C = P(C \mid G)/P(C \mid I)$. For a G-10 response, for example, $\Lambda_C=10$; for an I-10 response $\Lambda_C=1/10$; for an N response $\Lambda_C=1.0$. This letter-number pair response was used to prevent subjects from being confused by the ratios involved where $\Lambda_C>1$ means C favors "G" and $\Lambda_C<1$ means C favors "I".

This ZTD response is the result of a subject's holistic or global assessment of the probative value of an entire collection of evidence. Subjects made these responses following the thorough reading of the entire collection of evidence in each case. In this ZTD condition of our study, subjects made one such judgment for each of the twelve cases. The condition is called "zero task-decomposition" since the subjects performed the entire process of aggregating the evidence mentally and provided a single judgment indicating the probative force and direction of the evidence.

Partial Task-Decomposition (PTD) In this condition, subjects made exactly the same kinds of likelihood ratio estimates as in the ZTD condition, except that they made one such judgment for each item of "main-frame" testimony in each case. Such judgments were supported by the ancillary evidence in each block of evidence. On occasion, subjects were asked to refer to previous items of main-frame testimony in a case when there was some linkage between the items; for example, if the second of two items were contradictory with the first, subjects were asked to recall the first testimony in assessing the likelihood ratio for the second. Thus, if there were K items of "main-frame" testimony in a case, each subject made K assessments, one for each item.

Consider Case Cj which has some number K of evidence items. Bayes' rule prescribes a multiplicative procedure for combining the probative value for individual items to find the probative value of the entire case Cj. Thus, for case Cj, $\Lambda_{Cj} = \prod_{k=1}^{K} \tilde{\Lambda}_{jk}$, where $\tilde{\Lambda}_{jk}$ is the estimated likelihood ratio for the k^{th} item of main-frame testimony in case Cj. We can compare $\tilde{\Lambda}_{Cj}$ (ZTD), a person's estimated likelihood ratio in the zero task-decomposition condition, with $\Lambda_{\underset{\sim}{C}j}(PTD)$, a value calculated by simply multiplying together Λ_{jk} value for the individual testimonies in case Cj. Comparison between $\tilde{\Lambda}_{Cj}(ZTD)$ and $\Lambda_{Cj}(PTD)$ is one indication of internal consistency; are the probative value assessments of parts of a "case" consistent with the overall assessment of the entire case? This condition is called "partial task-decomposition," since the subject is now relieved of aggregating probativity assessments over the K evidence items in a case.

Complete Task Decomposition (CTD) In this condition, subjects' probativity-assessment tasks were decomposed to the "finest-grained" level of analysis allowed by our formal methods. This part of the study was conceptually the most difficult and required the most extensive instructions. As an illustration of the subjects' task in this condition, consider Equation 1 in the appendix. This shows the composition of a likelihood ratio, $\Lambda_{D_i^*}$, which describes the process of assessing the probative value of testimony D_i^* about event D, where D is circumstantial evidence of major facts-in-issue H_1 and H_2. Observe in Equation 1 that there are six conditional probability ingredients in $\Lambda_{D_i^*}$. Suppose D_i^*, from witness W_i, was an item of main-frame testimony in one of the cases the subjects evaluated. The subject's task was to estimate each of the six ingredient conditional probabilities required for a determination of $\Lambda_{D_i^*}$. For example, suppose that D represents the event that the defendant was at the scene/time of the crime. Subjects assessed the relative likelihood of this event under the assumption of guilt (H_1) and innocence (H_2) by means of the conditional probability ingredients $P(D|H_1)$ and $P(D|H_2)$. The remaining four ingredients in Equation 1 all concern the credibility of witness W_i, each of which the subject assessed. For example, $P(D_i^*|DH_1)$ asks: how probable is the witness testimony that defendant was at the scene/time, assuming that defendant was at the scene *and* guilty. The term $P(D_i^*|DH_2)$ asks: how probable is defendant's testimony, assuming defendant was at the scene *and* innocent. Subjects estimated such ingredient values on a probability scale [0,1] for Λ_{jk} values for each main-frame item of testimony in every case C_j.

In the actual estimation tasks in CTD, subjects responded to verbal descriptions of required conditional probabilities whose events corresponded to those in the case of concern. From these conditional probability estimates, we are able to calculate a value Λ_{jk} for any item of testimony in any case. These calculated values, for any subject, can be compared with that subject's estimate $\tilde{\Lambda}_{jk}$ for the same evidence item. In addition, for an entire case C_j we can determine Λ_{Cj} (CTD), a calculation of the probative value of entire case C_j based upon Λ_{ij} values calculated, in turn, from a subjects' conditional probability estimations.

In summary, our method allows three determinations of the probative value of each case for each subject; $\tilde{\Lambda}_{Cj}$ (ZTD), Λ_{Cj} (PTD), and Λ_{Cj} (CTD); these three values can be compared in consistency studies. In addition, our methods allow a comparison, for any subject and any evidence item k, between $\tilde{\Lambda}_{jk}$ (PTD) and calculated Λ_{jk} (CTD). Thus, we have consistency measures for whole-case probative value assessment and for individual evidence item assessment. As a final note, the three "levels" of task decomposition can be thought of as three sets of instructions of increasing specificity about the task or probativity assessment. The ZTD task leaves the entire aggregation and assessment burden on the subject. The PTD task requires holistic or global responses, but only to individual evidence items. The CTD task involves very specific instructions about the formally necessary linkages between events in reasoning chains established by foundation testimony. As we now relate, comparison of responses in these three conditions provides some interesting insight into human response characteristics in the task of weighing evidence.

V. EMPIRICAL STUDIES: RESULTS AND OBSERVATIONS

Following is a brief summary of major results obtained in those parts of our studies which concerned the assessment of the probative value of evidence. The accumulated data base is very large, and we have performed a variety of analyses on these data. A thorough account of the analyses and an extensive interpretation of the results are to be found in two recent research reports which, like others, are available to the reader interested in details (Schum and Martin, 1980c; 1981). Two types of results are presented here: those concerning the consistency of alternative ways of assessing the probative value of evidence for entire cases and for individual evidence items, and those bearing upon several interesting response patterns observed in the evaluation of several different species of evidence or inference structures. A few comments on our measurements and analyses are necessary before we begin.

Estimated or calculated values of likelihood ratio Λ are *vectors* having both probative *decision* and probative *force* properties. Probative direction specifies which major rival fact-in-issue (guilt or innocence) the evidence favors or "points toward" in an inferential sense. Probative force indicates the strength with which the evidence points toward the favored fact-in-issue. Some forms of statistical analyses are grossly misleading unless these two properties are examined separately. Consequently, we will talk about two "forms" of consistency. *Directional consistency* among two or more assessment methods means that the assessments in all methods agree in favoring the same fact-in-issue. *Force consistency* is measured by the degree to which two or more assessment methods assigned the same probative strength to the same evidence. For force consistency, we use a measure F, which indicates the factor by which two Λ values (estimated or calculated) differ. For example, if $\Lambda_1=10$ and $\Lambda_2=5$, then F=2. All measures of directional and force consistency are within-subject measures; that is, they compare two or more responses made by the same subject. One useful measure of between-subject consistency or agreement in probativity assessment is the familiar *concordance coefficient*. Applied to whole-case Λ_C it measures the extent to which the 20 subjects agree in rank-ordering the 12 cases in terms of their aggregate probative value. Applied to individual evidence items, this coefficient shows the agreement among the 20 subjects in rank-ordering the evidence items in a particular case in terms of their probative value.

Consistency Among Alternative Methods For Assessing The Probative Value of Evidence

Table 1 below contains a summary of results bearing upon the consistency of the three alternative response methods (ZTD, PTD, and CTD) and two other results of interest. Blank cells in Parts A and B simply indicate that the ZTD procedure produces no results for individual evidence items, since it involves a single estimate for an entire case; F is a pairwise measure. Part A shows directional consistency results for the four possible types of comparisons among ZTD, PTD, and CTD for whole-case Λ_C (Row A-1), and for individual evidence items Λ_{jk} (Row A-2). For example, 183 of the 240 possible Λ_C comparisons involving PTD and CTD were directionally consistent. Part B, rows B-1 and B-2 show probative force consistency measures F for whole-case Λ_C comparisons in

Table 1. General Consistency Results Summary

	ZTD PTD CTD	ZTD PTD	ZTD CTD	PTD CTD
A.				
1. Directional Consistency, Λ_C, 12 per subject	117/240 = 48.8%	132/240 = 55%	156/240 = 65%	183/240 = 76.3%
2. Directional Consistency, Λ_{jk}, 60 per subject	—	—	—	732/1200 = 61%
B.				
1. Force Consistency, Λ_C, Median F, directional agreement.	—	2.5* (132)	4.3 (156)	4.1 (183)
2. Force Consistency, Λ_C, Median F, directional disagreement.	—	10.0 (108)	30.0 (84)	20.0 (57)
3. Force Consistency, Λ_{jk}, Median F, directional agreement.	—	—	—	1.58 (729)
4. Force consistency, Λ_{jk}, Median F, directional disagreement.	—	—	—	2.30 (471)

C.	ZTD	PTD	CTD
1. Median Λ_C, (G)	2.0 (183)	5.0 (163)	10.0 (207)
2. Median Λ_C, (I)	4.0 (37)	5.0 (77)	8.0 (33)
3. Concordance: Ranking entire cases.	0.17	0.54	0.29

D. Concordance, individual items

Cases	1	2	3	4	5	6	7	8	9	10	11	12
PTD	.35	.25	.07	.18	.29	.60	.58	.24	.48	.32	.41	.56
CTD	.61	.60	.39	.55	.75	.61	.63	.53	.06	.61	.59	.39

* The lower the value, the greater the consistency.

which the assessments agreed directionally (B-1) and when they did not (B-2). Such separate analysis is necessary because F suppresses directionality. For example when Λ_C values in ZTD and PTD agreed in probative direction, they typically differed in probative force by a factor of 2.5. The number in parentheses under each F value indicates the number of comparisons over which a median F value was calculated. Rows B-3 and B-4 show median F for individual evidence item Λ_{jk} comparisons. For example, in the 729 instances in which PTD and CTD assessments of Λ_{jk} were directionally consistent, they differed typically by a factor of just 1.58.

Part C simply shows the typical (median) size of Λ_C estimated in ZTD or calculated in PTD or CTD; Row C-1 contains results when Λ_C favored guilt (G) and C-2 when Λ_C favored innocence I. Notice how Λ_C typically increases in size, whether it favors G or I, as the assessment task is decomposed to finer-grain levels. Row C-3 shows the concordance across the 20 subjects in rank-ordering the 12 cases in terms of Λ_C produced by each of the three methods; Λ_C in the PTD condition produces the greatest concordance among subjects. Finally, Part D shows the concordance or agreement among the 20 subjects in rank-ordering the evidence in each case using either PTD or CTD assessments. In all but two cases (9 and 12) the CTD procedure yielded Λ_{ij} values which were most in agreement across subjects.

ZTD In this response method, subjects had the task of assigning a single letter/number pair which indicated the probative strength and direction of the entire set of evidence in a "case." Though required to make only a simple response to each case, subjects had the complete burden of integrating or aggregating all of the evidence in a case. In determinations of the overall probative value of a case, ZTD fares worst in comparison with other methods. This method produces, pairwise, the weakest directional consistency with the other two methods, is most variable across subjects and across cases in directional comparisons, applies the weakest probative force, and has the lowest degree of concordance among subjects in ordering cases in terms of their probative value.

These results are not surprising. Similar results have been observed in other research on holistic assessments in comparison with other procedures (Edwards, *et al.*, 1968). There are several explanations for the typically weak force assessments provided by ZTD. The common explanation is the *misaggregation hypothesis*. This simply says that subjects, left to their own devices in combining large amounts of probabilistic evidence, use judgmental algorithms (or heuristics) which tend to let probative value inherent in evidence "leak out." By this hypothesis, we are all seen as "wasteful" processors of information. Another explanation is that the typically small Λ estimates used by subjects in ZTD simply reflects a *response bias* against using large numbers, particularly in situations in which there is ample evidence, even though conflicting, contradictory, and unreliable (Ducharme, 1970). Equally plausible is the notion that, in ZTD, subjects are free to discard any evidence for any reason. One could simply focus on a few "salient" features of evidence or only upon that evidence which agrees with prior expectations. This says that subjects are free to make the overall assessment task easier by reducing the number of items being considered. This strategy eliminates inconsistencies and reduces processing load; it may also yield weaker assessments if the subject is aware of the fact that evidence is being discarded.

PTD In this response method, subjects were forced to consider the probative value of the testimony of each major witness in a case. In this method Λ_C for an entire case is established by aggregating (multiplicatively) a subject's assessment of the probative value of each "main-frame" testimonial assertion. In such a method subjects are *partially* unburdened of aggregation, since they are never required to combine their assessments across other testimony. They do, however, have an increased response burden, being now required to provide one assessment for each of the "main-frame" items of testimony in a case. In directional consistency comparisons, PTD fares better than ZTD. There is a high directional consistency rate for PTD/CTD comparisons (A-1, A-2). These results are partially explainable by the fact that both use a common rule for aggregating probative value *across* the evidence items in a case (namely, multiplication). The closeness of these probative force results is also due to the apparent consistency of the ingredients subjects estimated in the CTD procedure, since a common between-item aggregation rule would not produce good agreement unless the ingredients in CTD were assessed in reasonable accordance with factors considered by subjects in the holistic PTD assessments. The PTD method produced the highest degree of concordance among subjects in ordering the 12 cases in terms of their probative value.

For individual evidence items, PTD agrees well with CTD in directional consistency and in force consistency (Part B). Concordance among subjects in ordering the evidence in individual cases is generally lower for PTD than for CTD. The PTD procedure, because it requires a focus on each individual witness, allows incorporation of factors which may be overlooked, discarded, or "integrated out" in the ZTD procedure. However, the PTD assessments for individual witnesses are holistic when compared to those assessments in CTD. Our results generally support the conjecture that additional features of probativity assessment required in CTD are overlooked, discarded, or "integrated out" in PTD.

CTD Most suprising to the authors was the overall consistency and adequacy of the many detailed probabilistic assessments made by subjects in the CTD procedure. Each subject made a total of 332 such assessments. In this response mode, subjects had a minimal aggregation burden but a maximal response or judgmental burden. In CTD, subjects were required to make judgments about the subtle linkages among events involved in the often-complex chains of reasoning from testimony to major facts-in-issue in each case. Judgments of the conditional probabilistic ingredients formally required in these chains, when aggregated by formally appropriate means, result in probativity assessments for entire cases and individual items which agree very well with assessments made by other, more holistic, means. This can only mean that the meticulous conditional probability assessments which CTD requires were performed in very reasonable and consistent ways by our subjects. As seen in Table 1-C, both directional and force consistency comparisons involving CTD are strongest for entire cases and for individual items. In addition, the degree of concordance among subjects in rank-ordering cases is second to PTD; in rank ordering individual items in a case, concordance among subjects is greatest using CTD procedures. CTD forces a person to look at the very fine-grained details of an inference task. Our study shows that persons required to bear the burden of such detailed analysis can perform the task in a manner very consistent with performance using other methods.

As the task of probativity assessment is decomposed into finer levels of analysis, a larger amount of the probative value latent in evidence is extracted and reflected in the assessment. This is explainable by at least two means. First, one locus of probative value in evidence is the possible conditional nonindependence of evidence items. Such nonindependence may easily be unrecognized or unaccounted for in ZTD. The PTD procedure alerts the subject to the existence of such nonindependence and simply tells the subject to account for it. In the CTD procedure, however, the method not only alerts the subject to such nonindependence, but also is instructive in how to reflect such nonindependence in assessments. Second and more obvious, as tasks are further decomposed, there is no chance that crucial evidence items, or factors concerning evidence items, will be overlooked, discarded, or "integrated out." For these reasons one expects more probative value in assessments in which more of the evidence in a collection is, in fact, incorporated and in which more of the subtle linkages among evidence items are reflected in the assessments.

We have also seen that there is greater concordance or agreement among individuals in probativity assessments as these tasks are further decomposed. The finer the level of decomposition, the more specific are the instructions required. It comes as no surprise to learn that agreement in judgment among persons is greater the more specific are the instructions about these judgments. We did, however, believe that the very specific conditional probabilistic judgments required in CTD would be difficult for our subjects to make. Results indicate that these judgments were made effectively, if not easily. The essence of task-decomposition in inference is that it forces consideration of finer-grained details of the task and ensures that such consideration is incorporated in overall assessments. Under procedures such as ZTD, where specific attention to these details is not enforced, entire evidence items or features of these items and their probabilistic linkages are easily discarded or overlooked by persons who may be unaware of their existence or who simply choose to ignore them.

Probative Assessment Characteristics For Different Inference Structures

Following is a collection of results obtained in a comparison of the PTD and CTD procedures for assessing the probative value of individual main-frame evidence items in each case. We first discuss results for evidence belonging to *simple* inference structures and then consider results for evidence in *complex* inference structures in which there were patterns of contradictory, corroboratively redundant, cumulatively redundant, or nonredundant evidence.

Simple Inference Structures These structures are characterized by the number of intermediate reasoning stages separating a single item of main-frame testimony and the ultimate facts-in-issue. A direct testimonial assertion about a major fact has no intermediate reasoning stages; therefore, its "level" of cascading is *zero*. The value of such testimony depends only on the witness's credibility. In our simple inference structures there were three other "levels" of cascading representing either one, two, or three intermediate reasoning stages. Evidence of opportunity typically may have either one or two intermediate reasoning stages, while evidence of means or motive typically have more as Wigmore noted (1937). So our essential results concern the effects, upon probative value assessment in the PTD and CTD condition, of the degree of logical remoteness between testimony and major facts-in-issue. Table 2-A below summarizes probative directional and force consistency between assessments in PTD and CTD.

Table 2. Directional and Force Consistency: PTD and CTD Conditions

	Logical Remoteness			
	0	1	2	3
A. Simple Inference Structures				
1) Directional Consistency	65% 104/160	73% 146/200	79% 253/320	73% 88/120
2) Force Consistency (Median F)	2.2	2.4	1.7	1.4
	CON	COR	CUR	NR
B. Complex Inference Structures				
1) Directional Consistency	41% 49/120	61% 49/80	65% 52/80	63% 76/120
2) Force Consistency (Median F)	2.0	2.1	1.9	2.0

Row A-1 shows that logical remoteness has little effect upon the proportion of *directionally* consistent assessment in PTD and CTD. Row A-2, however, shows that logical remoteness does influence the force consistency of these assessments; such consistency typically improves as the remoteness of testimony and facts-in-issue increases. The

https://doi.org/10.2307/3053534 Published online by Cambridge University Press

reason is quite apparent; assessed values of Λ_{jk} in PTD and those based upon subject ingredient estimates in CTD both decreased in size as the remoteness of testimony and major facts-in-issue increased. Abraham Lincoln's assessment of inference-upon-inference thus applies formally as well behaviorally; he is quoted as saying that inference-upon-inference frequently has the same strength as "soup made by boiling the shadow of a pigeon that has been starved to death" (Maguire *et al.*, 1973).

Complex Inference Structures "Complex" inference structures feature the testimony of more than one witness and involve various probabilistic linkages among events in the reasoning stages suggested by the testimony. All complex inference structures in our study involved two items of testimony, and inference structures were defined for various instances in which the second testimony was either contradictory (CON), corroboratively redundant (COR), cumulatively redundant (CUR), or nonredundant (NR) with the first item of testimony. Part B of Table 2 above contains results bearing upon the directional and force consistency of assessments of Λ_{jk} in PTD and CTD; the Λ_{jk} of concern in each case is for the *second* testimony in each structure, the one which either corroborates or is redundant or nonredundant with the first. As shown in Table 2, Part B, there is little difference in probative force consistency across complex inference structures. Directional consistency, however, is lower for contradictory structures than for the others. The explanation of this result requires a more detailed analysis, which we now present for each inference structure.

(1) Contradictory Testimony: Subject's probative response patterns to contradictory testimony in the PTD and CTD conditions are among the most interesting in this study. In the evidence evaluated by subjects there were six instances of contradictory testimony from two witnesses of apparently equal credibility; across 20 subjects we thus observed 120 pairs of PTD/CTD assessments relative to contradictory testimony. Of these 120 assessment pairs, 57 (48 percent) exhibited the following pattern. In the holistic PTD assessments, subjects either ignored the second and conflicting testimony (i.e. they assigned it no probative value) *or* they gave it value but made it agree directionally with the first item with which it was contradictory. Thus, in nearly half the contradictory evidence occasions subjects either ignored contradictory testimony or

treated such testimony as if it were corroborative. However, on these *same* occasions, their assessed ingredients in CTD for the Λ_{jk} values for contradictory testimony resulted in calculated Λ_{jk} values which were directionally the opposite of those for the first item of testimony. In short, asked to examine the fine-grain details of the tasks, subjects' assessments "brought out" the contradictory nature of evidence either overlooked or suppressed in their holistic PTD assessments.

This result brings to mind the so-called "primacy" effects others have observed in human holistic reactions to conflicting or contradictory evidence (e.g., Peterson and Ducharme, 1967; Pitz, 1969). The mind is set in motion in one direction by early evidence and somehow resists being moved in the other direction following later contradictory or conflicting evidence. In our study two items of contradictory testimony were always temporally adjacent. We have labeled the suppression of contradiction in such instances a "local" primacy effect to distinguish it from more global primacy effects which may persist over longer periods of time and over much intervening testimony (Schum, 1980b). The interesting result in our present study is that such primacy or contradiction-suppression is removed when subjects assess fine-grain logical details of evidence. It is important to note that such removal is not due to the aggregation models themselves, since calculated Λ_{jk} values depend entirely upon values of assessed ingredients. A subject's assessed ingredients, when coherently aggregated, "brought out" contradictions that this same person ignored or suppressed in holistic assessments. In making holistic or global assessments of any kind, individuals obviously resort to simplification strategies. Unfortunately, one such strategy may be the removal of contradiction or, worse yet, the incorporation of contradiction as if it were corroboration.

(2) Redundant Evidence Structures: We now examine the subjects' PTD and CTD response patterns to corroboratively or cumulatively redundant testimony. We have mentioned Lempert's (1977) concern about the extent to which jurors "double count" redundant testimony. Our results, taken seriously, suggest that his concern is certainly well founded. The systematic holistic PTD tendency to "double count" testimony from the second of two corroborative witnesses is shown in Table 3-A, and the holistic PTD tendency to overvalue cumulatively redundant testimony is shown in Table 3-B. Also shown in these tables is that such double-counting and over-evaluation tendencies do not appear in the CTD procedure for

probativity assessment. These results are best illustrated and summarized using the correlational statistics which we now describe.

Table 3. Subject Policies for Assessing Redundant Evidence in the PTD and CTD Conditions.

	PTD		CTD	
	r	b	r	b
A. Corroborative				
1. Instance 1	.96	.93	.72	.17
2. Instance 2	.97	1.07	.72	.23
3. Instance 3	.93	1.16	.65	.66
4. Instance 4	.83	.84	.42	.42
B. Cumulative				
1. Instance 1	.93	1.11	.34	.26
2. Instance 2	.79	.78	.25	.22
3. Instance 3	.73	1.00	.33	.27
4. Instance 4	.94	.99	.33	.19

In the case evidence subjects evaluated, there were four instances of testimony from two witnesses of nearly equal-appearing credibility who said the same thing. In Part A, for each of these four instances, are correlation coeffecients (r), and regression coefficients (b), in both PTD and CTD assessment conditions. To see how they were determined, take ZTD and instance 1 as an example. The value of r = 0.96 results from correlating, across all 20 subjects, the Λ_{jk} value assigned by a subject to the first testimony and the Λ_{jk} value assigned by the subject to the second (corroborating) testimony; values of b are found in the process. A value of r close to 1.0 means consistency across subjects in their policies for Λ assignment. The value b essentially tells what this policy was; if b = 1.0, this means that the subjects' values for the first and second items were identical; $b < 1$ means that Λ for the second item was smaller than Λ for the first; and $b>1$ means that Λ for the second item was greater than for the first.[8] So, for instance 1 in ZTD, the 20 subjects were nearly perfectly consistent in assigning the same probative weight to the second (corroborative) testimony as they assigned to the first. As can be seen, essentially the same thing happens in all four instances; in PTD (holistic assessment) subjects systematically double-counted corroboratively redundant testimony.

[8] These stated policies assume that the calculated intercepts are zero or near zero, which they were in all but one condition in Part A. An intercept of zero says that subject's Λ values have no initial bias toward guilt or innocence.

In the CTD condition the generally lower r values mean less consistency in policy; this is to be expected given the very large number of conditional probability ingredients involved in determining the Λ_{jk} value being correlated. Notice however that b is much less than 1.0 in every instance; this means that the probative weight assigned to the corroboratively redundant testimony was typically smaller than the weight assigned in the original testimony. In short, subjects' CTD estimates, when formally aggregated, "bring out" the redundance apparently overlooked in the holistic PTD assessments.

Part B of Table 3 shows the same analysis for the four instances of cumulatively redundant testimony in our evidence. Essentially the same results occur across subjects though it is evident that there is less consistency in estimating Λ_{jk} ingredients than there is in holistic estimates of Λ_{jk}. Cumulative redundance is a more subtle effect than is corroborative redundance, and Λ_{jk} models for cumulatively redundant testimony have many more ingredients than those for corroboratively redundant testimony. Nevertheless, cumulative redundancy overlooked in PTD is "brought out" in the fine-grained analysis in CTD. Once again we emphasize that the "bringing out" in CTD of subtleties due to contradiction or redundance are not necessarily due to the formal aggregation models for Λ_{jk}; it is the subject-assessed ingredients which allow these calculations to bring out these subtleties.

(3) Nonredundant Structures: In the evidence that subjects evaluated there were six instances in which the first of two testimonial assertions was contrived to be probatively facilitative on the second. Apparently, our contrived evidence failed to appear facilitative since, in both PTD and CTD, the second testimony was weighted essentially the same as the first. The formal distinction between redundant and nonredundant testimonial evidence is subtle and involves no sharply defined boundary. In fact, measures of the redundance of evidence suggest a continuum along which evidence may be placed in terms of whether it is redundant or facilitative (Schum, 1979b).

VI. CONCLUSION

The formal research summarized in this paper provides examples of the method whereby various identifiable forms of evidence can be subjected to a fine-grained analysis in which the subtleties in evidence can be identified and systematically examined. The stimulus and guidance for such research have come in large measure from existing evidence scholarship in jurisprudence, there being no similar body of scholarship to be found in other areas in which reasoning from inconclusive evidence is commonly encountered. Our formal research has concerned some matters of which jurists are already aware, such as the complex interplay between witness credibility and the value of what the witness says in a determination of the probative value of testimony by the witness. In fact, some of our specific formal studies on such areas as testimonial redundance have been enhanced by careful examination of evidentiary distinctions which jurists commonly recognize or, in a few instances, fail to recognize. It is generally the case that difficulty in the formal analysis of some forms of evidence (e.g. hearsay, cumulatively redundant evidence) parallels the difficulty jurists experience in formulating specific prescriptions concerning the relevance and admissibility of such evidence.

Our formal research also concerns matters about which there is only infrequent or oblique reference in evidence scholarship—event rarity and transitivity issues being examples. One distinct virtue of the formal analysis of various evidentiary patterns is that it allows one to observe the way in which a variety of evidence-related formal systems can be brought to bear in the study and analysis of evidence subleties. Thus, we are able to show how the theory of "signal detection" provides very useful concepts and methods in the study of credibility-related ingredients in the analysis of the probative value of testimonial evidence. Similarly, concepts from "statistical communication theory" (or "information theory") allow one to be more precise in formulating problems relating to possible redundance in certain evidence.

How exciting or informative are the results of the empirical studies we summarize depends to some extent on the reader's expectations. Some of our results are very similar to those found in other, more abstract, laboratory studies; other results are unique because of the complexity of the tasks our research subjects performed. In these concluding comments about our empirical results we do *not* wish to leave the reader with the impression that these studies are simply further examples of psychological research demonstrating the inadequacy of human performance on inference-related tasks. In fact, our results show how well persons who are given adequate instructions about reasonably well-formulated problems can respond to a variety of subtle aspects of evidence and reflect these subtleties in their responses. Demonstrating human inadequacy at an inferential task presupposes uncontroversial performance standards, well-posed tasks, and adequate instructions. *No* empirical study of inference (including ours) has the first, and regrettably few have the other two.

A general result of our study is that individuals can capture in their probativity-assessment responses an assortment of subtle aspects of evidence provided that their probativity-assessment tasks are decomposed to a level at which these subtleties are exposed. Further, the concordance or agreement among the probativity assessors in our study was highest for decomposed assessment tasks. Finally, global, holistic, or nondecomposed assessments of the probative value in a collection of evidence are the most variable across persons and frequently disagree directionally with assessments made by other methods involving task decomposition. One suggestion is that individuals asked to mentally aggregate a large collection of evidence may ignore, discard, or integrate over contradictory evidence and otherwise overlook other subtleties in evidence. Our message to jurists cannot, for obvious reasons, be that fact finders' tasks ought to be decomposed in the manner in which they were in our study. Presumably, the deliberation process following a trial encourages a fact finder to consider factors which others have noted but which he/she has discarded or ignored; such mutual enlightenment, however, cannot be guaranteed.

Perhaps the most striking results of our study concern the manner in which our research subjects assessed the value of contradictory and of redundant testimony. Quite startling is the frequently-observed holistic tendency to make contradictory testimony either probatively valueless or, what seems worse, corroborative; such behavior, however, is certainly not unheard-of in more abstract studies of human inference. Our studies show the existence of local as well as global "primacy" effects in which, apparently, the mind resists changes in the direction of opinion revisions. The most systematic result in our study concerns the holistic tendency to "double count" corroboratively redundant testimony. Neither tendency is apparent when subjects are allowed to examine and respond to the fine-grained details of evidence having these characteristics.

If, as asserted earlier, the fact finder's task cannot be decomposed and the post-trial deliberation process cannot guarantee appropriate assessment of various forms of evidence, then it is left to the skill of one counsel in "decomposing" the arguments made by the other. Our formal research strongly suggests the essential formal adequacy of the rules and procedures for this process of "beating and boulting out the truth" (Hale, 1739). Our empirical studies suggest that attentive fact finders with reasonable intellectual skills can incorporate the many subtleties in evidence if, at least, they are alerted to the existence of these subtleties. Thus, the double-counting of redundant testimony and the "local primacy" of earlier testimony that is contradicted later are examples of common reasoning inconsistencies exhibited by many people which can, perhaps, be overcome by an equally attentive counsel.

TECHNICAL APPENDIX

A. THREE EXAMPLES OF LIKELIHOOD RATIO Λ DETERMINATION FOR THE THREE INFERENCE STRUCTURES SHOWN IN FIGURE 1 IN THE TEXT

Figure 1-A in the text shows the simplest possible case of cascaded inference. Witness W_i testifies that event D occurred; event D is circumstantial evidence bearing on major rival facts-in-issue H_1 and H_2. As a fact finder, what we have is W_i's testimonial assertion D_i^* that event D occurred. D_i^* and D are not the same events, and we shall be misled if we treat them so. The reason is that, unless we believe W_i to be perfectly credible, testimony D_i^* is consistent both with D, the actual occurrence of the event, and D^C, the nonoccurrence of this event. In this case, what can condition or change our opinion about the relative likelihood of H_1 and H_2 is the event D_i^* representing the testimony of W_i. In this inference structure the likelihood ratio for testimony D_i^* is as follows:

$$\Lambda_{D_i^*} = \frac{P(D_i^*|H_1)}{P(D_i^*|H_2)} = \frac{P(D|H_1)[P(D_i^*|DH_1) - P(D_i^*|D^CH_1)] + P(D_i^*|D^CH_1)}{P(D|H_2)[P(D_i^*|DH_2) - P(D_i^*|D^CH_2)] + P(D_i^*|D^CH_2)} \quad (1)$$

In this expression the conditional probabilities $P(D|H_1)$ and $P(D|H_2)$ express the strength of the linkage between D and H_1 and H_2; in words, these probabilities prescribe how probatively valuable is event D. They indicate the strength of the linkage between D and H_1, H_2. All other terms concern the credibility of witness W_i and refer to the

https://doi.org/10.2307/3053534 Published online by Cambridge University Press

linkage between testimony D_i^* and D, the matter asserted. So far, intuition is supported by $\Lambda_{D_i^*}$; the value of an item of testimony depends upon the importance of what the source has to say and upon the credibility of the source.

There are two additional features of $\Lambda_{D_i^*}$ which we must also notice. Observe that the terms $P(D|H_1)$ and $P(D|H_2)$ occur separately in Equation 1 and not together as the ratio $P(D|H_1/P(D|H_2)$. This fact tells us that the probative value of testimony also depends upon the *rarity* of D, the event being reported. The reason is that the ratio of two numbers suppresses information about the precise values of the numbers; e.g., $4=0.40/0.10 = 0.04/0.01$. In $\Lambda_{D_i^*}$ we must have the precise values of $P(D|H_1)$ and $P(D|H_2)$ and not simply their ratio; such precision preserves the rarity of events, to which Λ is sensitive. In general, the rarer the event reported, the stronger credibility we may require from the source of information about the source.

The other feature concerns the four credibility-related ingredients of $\Lambda_{D_i^*}$ in Equation 1. Take $P(D_i^*|DH_1)$ for example. By itself, the conditional probability $P(D_i^*|D)$ is called a "hit" probability or a "true positive," and it expresses how likely testimony D_i^* is if event D actually occurred. The addition of the conditioning term H_1 in $P(D_i^*|DH_1)$ tells us essentially that W_i's "hit probability" may depend upon H_1. A similar process applies to another ingredient $P(D_i^*|D^CH_1)$, which is called "false positive," and also to these hit and false positives when H_2 is true. Here is the essential message; in assessing the probative value of testimony D_i^* we must consider whether or not the likelihood of testimony D_i^* depends upon factors other than the occurrence or nonoccurrence of the events being reported. Our formal process makes clear that the observational and reporting behavior of a witness may contain probative value over and above the probative value of the event being reported. In short, the witness' behavior can be probative in a number of ways. The manner in which credibility-related ingredients occur in Λ expressions allows us to incorporate a wide variety of subtleties associated with the behavior of witnesses including their observational sensitivity and many motivational considerations.

Figure 1-B represents a situation in which witness W_j offers testimony E_j^* that event E occurred; the occurrence of E is circumstantial on D which, in turn, is circumstantial

on major facts-in-issue H_1 and H_2. Equation 2 shows $\Lambda_{E_j^*}$ in a special case which we shall identify:

$$\Lambda_{E_j^*} = \frac{P(E_j^*|H_1)}{P(E_j^*|H_2)}$$

$$= \frac{P(D|H_1)[P(E|DH_1) - P(E|D^CH_1)] + P(E|D^CH_1) + \left[\frac{P(E_j^*|E)}{P(E_j^*|E^C)} - 1\right]^{-1}}{P(D|H_2)[P(E|DH_2) - P(E|D^CH_2)] + P(E|D^CH_2) + \left[\frac{P(E_j^*|E)}{P(E_j^*|E^C)} - 1\right]^{-1}} \quad (2)$$

This structure reveals three reasoning stages: from testimony E_j^* to events E, E^C; from events E, E^C to events D, D^C; and from events D, D^C, to ultimate facts-in-issue H_1, H_2. Examination of Equation 2 shows probative ingredients for each stage of reasoning. The "special-case" nature of this expression concerns the foundation reasoning stage from testimony E_j^* to events E, E^C. In the more general expression for $\Lambda_{E_j^*}$, the hit and false-positive $P(E_j^*|E)$ and $P(E_j^*|E^C)$ have other conditioning terms, namely, the four possible combinations of one of D, D^C and H_1, H_2. The special case arises when one assumes that these hit and false positives do not depend upon any events "higher" in the chain of reasoning. The elimination of these higher-order conditioning events is accomplished by conditional independence assumptions. For example, $P(E_j^*|EDH_1) = P(E_j^*|E)$ is the assumption that testimony E_j^* is independent of D and H_1, given event E. In short, this is an assumption that the credibility-related hit probability $P(E_j^*|E)$ does not depend upon other events in the reasoning chain.

Figure 1-C shows a case in which two witnesses W_i and W_j both testify that event D occurred. Here is an instance in which the probative value of one item of testimony depends upon previous evidence; our Λ formulations account for such dependency, allowing us to represent a variety of subtle effects of one evidence item on another. This example allows us to show the degree to which testimony D_j^* is probatively redundant, since W_j reports the *same* event as did earlier Witness W_i.

The probative value of testimony from the first witness W_i is given by Equation 1. Now consider W_j who also testifies that D occurred. We must now examine testimony D_j^* *in light of* prior testimony D_i^*, since there is an obvious logical relationship—namely, both witnesses say the same thing. Formally, the likelihood ratio for D_j^*, given D_i^*, is prescribed by:

$$\Lambda D_j^*|D_i^* = \frac{P(D_j^*|H_1D_i^*)}{P(D_j^*|H_2D_i^*)}$$

$$= \frac{P(D|D_i^*H_1)[P(D_j^*|DD_i^*H_1) - P(D_j^*|D^CD_i^*H_1)] + P(D_j^*|D^CD_i^*H_1)}{P(D|D_i^*H_2)[P(D_j^*|DD_i^*H_2) - P(D_j^*|D^CD_i^*H_2)] + P(D_j^*|D^CD_i^*H_2)} \quad (3)$$

This formalization shows that there are two interesting factors in the relationship between D_i^* and D_j^*. The first concerns the terms $P(D|D_i^*H_1)$ and $P(D|D_i^*H_2)$. Basically, these terms prescribe the "residual" probative value in D remaining after W_i's testimony; how much is left depends upon W_i's credibility. If W_i is perfectly credible, then there is no probativity left for the testimony of W_j; if you believe W_i, then testimony by W_j should tell you nothing more in inference about H_1 and H_2. The other factor concerns the possible conditioning of testimony D_j^* by testimony D_i^*. Essentially, this allows for the incorporation of factors associated with possible influence of one witness on another. We may believe, for example, that W_i told W_j what to testify. If so, there is room in Equation 3 for incorporating such effects and adjusting probative values accordingly.

B. MEASURES OF EVENT REDUNDANCY

For the *cumulative* case, we define the redundancy of event F, knowing event E, as:

$$R_{cum} = 1 - \frac{\text{Log}^{L_{F|E}}}{\text{Log } L_F},$$

where L_F is a likelihood ratio measure of the probativity of event F on facts-in-issue, and $L_{F|E}$ is a likelihood ratio measure of the probativity of event F in light of event E. If the occurrence of E makes F highly probable under both facts-in-issue, then $L_{F|E} = 1.0$. Since Log 1 = zero, this makes $R_{cum} = 1.0$, it maximum value. If knowing E causes no change in the probativity of F on facts-in-issue, then $L_{F|E} = L_F$ and so R_{cum} = zero; this means that F is not probatively redundant if you knew that E occurred. So, R_{cum} is a number between zero and one which indicates the extent of redundance in event F if you also knew that event E occurred.

In the *corroborative* case, we define:

$$R_{cor} = 1 - \frac{\text{Log } L_{E|E}}{\text{Log } L_E}, \quad (5)$$

and note immediately that R_{cor} *must always equal 1.0*. The reason, of course, is that the second discovery of the same event cannot be probative, $L_{E|E} = 1.0$. Since Log 1.0 = 0, R_{cor} always equals one. In other words, event E is always perfectly redundant with itself. This is why we have said in the text that corroborative redundance is a special case of cumulative redundance.

REFERENCES

CLEARY, Edward (1972) *McCormick on Evidence*. St Paul.: West.

COHEN, L. Jonathan (1977) *The Probable and the Provable*. Oxford: Clarendon.

——— (1981) "Subjective Probability and the Paradox of the Gatecrasher," 1981 *Arizona State Law Journal* 627.

DUCHARME, Wesley M. (1970) "A Response Bias Explanation of Conservative Inference," 85 *Journal of Experimental Psychology* 66.

EDWARDS, Ward D. (1962) "Dynamic Decision Theory and Probabilistic Information Processing," 4(2) *Human Factors* 59.

EDWARDS, Ward D., William L. HAYS, and Barabara C. GOODMAN(1968) "Probabilistic Information Processing Systems: Design and Evaluation," 4 *IEEE Transactions on Systems Science and Cybernetics* 248.

EGAN, James P. (1975) *Signal Detection Theory and ROC Analysis*. N.Y.: Academic Press.

EGGLESTON, Richard (1978) *Evidence, Proof and Probability*. London: Weidenfeld and Nicolson.

EILS, I., David SEAVER, and Ward EDWARDS (1977) Developing a Technology of Probabilistic Inference: Aggregating by Averaging Reduces Conservation. 77 Social Science Research Institute, University of Southern California, Los Angeles.

FINKELSTEIN, Michael O. and William B. FAIRLEY (1970) "A Bayesian Approach to Identification Evidence," 83 *Harvard Law Review* 489.

HALE, Sir Matthew (1739) *The History of the Common Law of England*. Chicago: University of Chicago Press (1971 reprint).

HUME, David (1888) *Treatise of Human Nature*, Book 1, Part 3, Section 8. Oxford: Clarendon Press.

KAYE, David (1979) "The Paradox of the Gatecrasher and Other Stories," 1979 *Arizona State Law Journal* 101.

——— (1981) "Paradoxes, Gedanken Experiments and the Burden of Proof: A Response to Dr. Cohen's Reply," 1981 *Arizona State Law Journal* 635.

KEYSER, Cassius J. (1929) "On the Study of Legal Science," 1929 *Yale Law Journal* 38.

LEMPERT, Richard O. (1977) "Modeling Revelance," 75 *Michigan Law Review* 1021.

LEMPERT, Richard O. and Stephen A. SALTZBURG (1977) *A Modern Approach to Evidence*. St. Paul: West.

MAGUIRE, John M., J. WEINSTEIN, J. CHADBOURN, and J. MANSFIELD (1973) *Evidence: Cases and Materials on Evidence*. Mineola, N.Y.: Foundation Press.

MARTIN, Anne W. (1979) Cascaded Inference and Hearsay. Rice University Department of Psychology Research Report 79-03. Houston Texas.

——— (1980) A General Algorithm for Determining Likelihood Ratios in Cascaded Inference. Rice University Department of Psychology Research Report 80-03, Houston Texas.

PETERSON, Cameron R. and Wesley M. DUCHARME (1967) "A Primacy Effect in Subjective Probability Revision," 73 *Journal of Experimental Psychology* 61.

PITZ, Gordon F. (1969) "An Inertia Effect (Resistance to Change) in the Revision of Opinion," 23 *Canadian Journal of Psychology* 24.

RAPOPORT, Ammon and Thomas WALLSTEIN (1972) "Individual Decision Behavior," 23 *Annual Review of Psychology*, 131.

SAKS, Michael J., and Robert F. KIDD (1981) "Human Information Processing and Adjudication: Trial by Heuristics," 15 *Law & Society Review* 123.

SCHAFER, G. (1976) *A Mathematical Theory of Evidence*. Princeton N.J.: Princeton University Press.

SCHUM, David A. (1977a) "The Behavioral Richness of Cascaded Inference Models: Examples in Jurisprudence," in Castellan, Pisoni, and Potts (eds.), *Cognitive Theory* (Vol. II). Hillsdale, N.J.: L. Erlbaum Press.

——— (1977b) "Contrast Effects in Inference: On the Conditioning of Current Evidence by Prior Evidence," 18 *Organizational Behavior and Human Performance* 217.

——— (1979a) "A Review of a Case Against Blaise Pascal and his Heirs," 77 *University of Michigan Law Review* 446.

——— (1979b) On Factors Which Influence the Redundancy of Cumulative and Corroborative Testimonial Evidence. Rice University Psychology Dept. Research Report 79-02, Houston, Texas.

——— (1979c) A Bayesian Account of Transitivity and Other Order - Related Effects In Chains of Inferential Reasoning. Rice University Psychology Department Research Report 79-04, Houston, Texas.

——— (1980) "Current Developments in Research on Cascaded Inference," in Thomas S. Wallstein (ed.), *Cognitive Processes in Decision and Choice Behavior*. Hillsdale, N.J.: L. Erlbaum Press.

——— (1981a) "Sorting out the Effects of Witness Sensitivity and Response Criterion Placement Upon the Inferential Value of Testimonial Evidence," 27 *Organizational Behavior and Human Performance* 153.

——— (1981b) Assessing the Probative Value of Equivocal Testimony Or No Testimony on a Relevant Issue at Trial. Rice University Department of Psychology Research Report #81-04, Houston, Texas.

——— (1981c) Formalizing the Process of Assessing the Probative Value of Alibi Testimony. Rice University Psychology Dept. Research Report 81-05, Houston, Texas.

SCHUM, David A. and Wesley M. DUCHARME (1971) "Comments on the Relationship between the Impact and the Reliability of Evidence," 6 *Organizational Behavior and Human Performance* 111.

SCHUM, David A. and Clinton, W. KELLEY (1973) "A Problem in Cascaded Inference: Determining the Inferential Impact of Confirming and Conflicting Reports from Several Unreliable Sources," 10 *Organizational Behavior and Human Performance* 404.

SCHUM, David A. and Anne W. MARTIN (1980a) Empirical Studies of Cascaded Inference in Jurisprudence: Methodological Considerations. Rice University Department of Psychology Research Report 80-01, Houston, Texas.

——— (1980b) Probabilistic Opinion Revision on the Basis of Evidence at Trial: A Baconian or a Pascalian Process? Rice University Department of Psychology Research Report 80-02, Houston, Texas.

——— (1980c) On the Internal Consistency of Assessments of the Probative Value of Evidence. Rice University Dept. of Psychology Research Report 80-04, Houston, Texas.

——— (1981) Assessing the Probative Value of Evidence in Various Inference Structures. Rice University Dept. of Psychology Research Report 81-02, Houston, Texas.

SLOVIC, Paul and Sarah LICHTENSTEIN (1971) "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment," *6 Organizational Behavior and Human Performance* 649.

STANILAND, Alan C. (1966) *Patterns of Redundancy*. Cambridge: The University Press.

STEVENS, Stanley S. (1975) *Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects*. N.Y.: Wiley and Sons.

SWETS, John A. (1964) *Signal Detection and Recognition by Human Observers: Contemporary Readings*. N.Y.: Wiley and Sons.

TRIBE, Laurence (1971) "Precision and Ritual in the Legal Process," 84 *Harvard Law Review* 1810.

——— (1974) "Triangulating Hearsay," 87 *Harvard Law Review* 957.

WIGMORE, John H. (1937) *The Science of Judicial Proof as Given by Logic, Psychology, and General Experience, and Illustrated in Judicial Trials* (3rd ed.). Boston: Little, Brown.

WINKLER, Robert and Allen MURPHY (1973) "Experiments in the Laboratory and the Real World," 10 *Organizational Behavior and Human Performance* 252.

ZADEH, Lofti (1978) "Fuzzy Sets As a Basis For a Theory of Possibility," 1 *Fuzzy Sets And Systems*.