### Inside the juror

# The psychology of juror decision making

Edited by Reid Hastie University of Colorado



Published by the Press Syndicate of the University of Cambridge The Pitt Building, Trumpington Street, Cambridge CB2 1RP 40 West 20th Street, New York, NY 10011-4211, USA 10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1993

First published 1993 First paperback edition 1994

#### Library of Congress Cataloging-in-Publication Data

Inside the juror: the psychology of juror decision making / edited by Reid Hastie. p. cm. - (Cambridge series on judgment and decision making) ISBN 0-521-41988-3 (hc) 1. Jury - United States - Decision making. 2. Jury - United States -Psychological aspects. I. Hastie, Reid. II. Series. KF8972.157 1993 347.73'0752 - dc20 [347.307752] 92-5018 CIP

A catalog record for this book is available from the British Library.

ISBN 0-521-41988-3 hardback ISBN 0-521-47755-7 paperback

Transferred to digital printing 2004

### Contents

Ser	ies preface	page vii
	t of contributors	viii ·
Eat	itor s prejace	1X
ACI	knowleagments	x
Paı	t I: Models of juror decision making	
1	Introduction	3
	Reid Hastie	
2	Some steps between attitudes and verdicts <i>Phoebe C. Ellsworth</i>	42
3	The influence of outcome information and attitudes on	
-	juror decision making in search and seizure cases	65
	Jonathan D. Casper and Kennette M. Benedict	
4	Algebraic models of juror decision processes	84
	Reid Hastie	
5	Stochastic models of juror decision making	116
	Norbert Kerr	
6	Formal and empirical research on cascaded inference in	
	jurisprudence	136
	David A. Schum and Anne W. Martin	
7	Argument structuring and evidence evaluation	175
	David A. Schum	
8	The story model for juror decision making	192
	Nancy Pennington and Reid Hastie	
Pai	t II: Commentaries	
9	Notes on the sampling of stimulus cases and the	
	measurement of responses in research on juror	
	decision making	225
	Robyn M. Dawes	
10	Sausages and the law: Juror decisions in the much larger	
	justice system	229
	Joseph B. Kadane	

vi Contents

11	A rational game theory framework for the analysis of legal and criminal decision making <i>Fhud Kalai</i>	235
12	Why do jury research?	242
	Richard O. Lempert	
13	Two conceptions of the juror	255
	Lola Lopes	
14	A mathematician comments on models of juror decision making	263
	Sandy Zabell	200
Ind Ind	ex of names ex of subjects	271 276

## 8 The story model for juror decision making

#### Nancy Pennington and Reid Hastie

#### Introduction

The goal of our research over the past ten years has been to understand the cognitive strategies that individual jurors use to process trial information in order to make a decision prior to deliberation. We have approached this goal with the perspective of psychologists who are interested in how people think and behave. First, we have developed a theory that we believe describes the cognitive strategies that jurors use. We call this theory the story model, and it is described in the first section of the paper. Second, we have conducted extensive empirical work to test the theory. This work is summarized in the second section of the paper. The story model has been developed in the context of criminal trials, so it will be presented and discussed in those terms. In the final section of the paper, we discuss some of our current research directions.

#### The story model

We call our theory the story model because we propose that a central cognitive process in juror decision making is *story construction* (Bennett & Feldman, 1981; Pennington, 1981, 1991; Pennington & Hastie, 1980, 1981a, 1981b, 1986, 1988, 1992). Although story construction is central in our theory and has been the focus of most of our empirical research, it is but one of three processes that we propose. In overview, the story model includes three component processes: (A) evidence evaluation through story construction, (B) representation of the decision alternatives by learning verdict category attributes, and (C) reaching a decision through the classification of the story

Research reported in this paper was supported by the National Science Foundation, Law and Social Sciences Program.



Figure 8.1. The story model for juror decision making.

into the best fitting verdict category (see Figure 8.1). In addition to descriptions of processing stages, one central claim of the model is that the story the juror constructs *determines* the juror's decision. As part of the theory, we also propose four certainty principles – coverage, coherence, uniqueness, and goodness-of-fit – that govern which story will be accepted, which decision will be selected, and the confidence or level of certainty with which a particular decision will be made.

In the next sections of the paper we describe the processing stages proposed in the story model and the certainty principles that govern them. In order to illustrate our ideas with examples, we will draw on one of the simulated trials that we have used in our research, *Commonwealth of Massachusetts v. Johnson*. In this trial, the defendant Frank Johnson is charged with first-degree murder. The undisputed background events include the following: the defendant Johnson, and the victim, Alan Caldwell, had a quarrel early on the day of Caldwell's death. At that time, Caldwell threatened Johnson with a razor. Later in the evening, they were again at the same bar. They went outside together, got into a fight, and Johnson knifed Caldwell, resulting in Caldwell's death. The events under dispute include whether or not Caldwell pulled a razor in the evening fight, whether Johnson actively stabbed Caldwell or merely held his knife out to protect himself, how they got outside together, whether or not Johnson intentionally went home and got his knife, whether Johnson went back to the bar to find Caldwell or went to the bar because it was his habit, etc.<sup>1</sup>

#### Constructing a story

The story model is based on the hypothesis that jurors *impose* a narrative story organization on trial information. According to the theory, the story will be constructed from three types of knowledge (see Figure 8.1 top left): (A) case-specific information acquired during the trial (e.g., statements made by witnesses about past events relevant to the decision); (B) knowledge about events similar in content to those that are the topic of dispute (e.g., knowledge about a similar crime in the juror's community); and (C) generic expectations about what makes a complete story (e.g., knowledge that human actions are usually motivated by goals). This constructive mental activity results in one or more *interpretations* of the evidence that have a narrative story form (Figure 8.1, top right). One of these interpretations (stories) will be accepted by the juror as the best explanation of the evidence. The story that is accepted is the one that provides the greatest coverage of the evidence and is the most coherent, as determined by the particular juror.

Active story construction. When we hypothesize that jurors *impose* a narrative organization on evidence, we mean that jurors engage in an active, constructive comprehension process in which evidence is organized, elaborated, and interpreted by them during the course of the trial. In part, this activity occurs because comprehension is inherently a constructive process for even the simplest discourse (Collins, Brown, & Larkin, 1980; Crothers, 1979; Kintsch, 1974, 1988). To illustrate this point in general, suppose a listener is told a simple narrative, "Billy went to Johnny's birthday party. When all the children were there, Johnny opened his presents. Later, they sang Happy Birthday and Johnny blew out the candles." Many listeners will infer spontaneously, and most will agree when asked, that there was a cake at the birthday party. Yet, no cake is mentioned in the sentences above; indeed it is not certain that there was a cake. The cake is inferred

<sup>&</sup>lt;sup>1</sup> This trial has been used extensively in mock jury research (Hastie, Penrod, & Pennington, 1983; Ellsworth, 1988) and has been judged by experienced attorneys and trial judges to be a typical felony trial; see Appendix to Chapter 1 of this book for more details.

because we share knowledge about birthday party traditions and about the physical world (the candles had to be on something). Another illustration comes with the comprehension of the sentence, "The policeman held up his hand and stopped the car." Most of us understand this sentence in the cultural context of the policeman's authority, shared signals, a driver watching the policeman but controlling the car, etc. Indeed, this is a sentence that would be puzzling to a person from a different culture.

The constructive nature of comprehension is especially relevant in the context of legal trials in which characteristics of the trial evidence make it unwieldy and unstory-like. First there is a lot of evidence, often presented over a duration of several days. Second, evidence presentation typically appears in a disconnected question and answer format; different witnesses testify to different pieces of the chain of events, usually not in temporal or causal order; and witnesses are typically not allowed to speculate on necessary connecting events such as why certain actions were carried out, or what emotional reaction a person had to a certain event. The attorney's opportunity to remedy the unstory-like form of evidence presentation occurs during the presentation of opening and/or closing arguments, but this opportunity is not always taken.

According to the story model, stories are constructed by reasoning from world knowledge and from evidence. Some potential story elements are accepted as true directly on the basis of their appearance as evidence from one or more credible sources; they are reasonably well established as fact. Which of these events will appear as relevant depends on the interpretation assigned to the fact from its causal relatedness to other events. The inclusion in the story of other evidence, inferred events, and causal relations between them is the result of a wide variety of deductive and inductive reasoning procedures applied to the evidence and world knowledge (Collins, 1978; Collins & Michalski, 1989).

Analyses of inference chains leading to story events reveal that intermediate conclusions are established by converging lines of reasoning that rely on deduction from world knowledge, analogies to experienced and hypothetical episodes, and reasoning by contradiction (Pennington, 1991; Pennington & Hastie, 1980). A typical deduction from world knowledge in the "Johnson case" consists of the following premise (P1 – P3) and conclusion (C) structure:

- P1. A person who is big and known to be a troublemaker causes people to be afraid.
- P2. Caldwell was big.
- P3. Caldwell was known to be a troublemaker.
- C. Johnson was afraid.

In this example, the juror matches features of Caldwell from undisputed evidence (P2) and a previous inferential conclusion (P3) to world knowledge about the consequences of being confronted with such a person (P1) to infer that Johnson was afraid (C).<sup>2</sup>

Confidence in the conclusion of an inference is assessed by reasoning by analogy to other experiences and by evaluating alternate conclusions that would contradict the initial conclusion. For example, the same juror who provided the premise-conclusion example just mentioned, continued with, "If someone like Caldwell came up to me in a bar and threatened me, I would be afraid." Alternate reactions were also considered, "I don't think Johnson was angry. If he had been angry, he would have gone right back to the bar. He didn't go right back." This alternative is rejected, "No, Johnson was afraid of Caldwell and he took his knife with him because he was afraid."

Different jurors will construct different stories, and a central claim of the theory is that the story will determine the decision that the juror reaches. Because all jurors hear the same evidence, and have the same general knowledge about the expected structure of stories, differences in story construction must arise from differences in world knowledge, that is, differences in experiences and beliefs about the social world. In contrast to the example inference above, another juror might believe that confrontations by bullies are a challenge to manly pride and that as a result, anger is the more likely response. This particular inference may be a keystone in an evolving interpretation of the evidence that is completely different from that of the previous juror.

The structure of stories. Stories involve human action sequences connected by relationships of physical causality and intentional causality between events. In its loosest form, a story could be described as a "causal chain" of events in which events are connected by causal relationships of necessity and sufficiency (Trabasso & van den Broek, 1985). However, psychological research on discourse comprehension suggests that story causal chains have additional higher order structure both when considering the discourse itself and when considering the listener or reader's "mental representations"

<sup>2</sup> It is the certainty of the conclusion C as a function of the levels of certainty of P1, P2, and P3, and the strengths of the relationships between the premises and conclusion that probabilistic (and heuristic) theories of inference were designed to model. It is at this point that Bayesian or fuzzy set calculations could be incorporated into the story model to yield the level of certainty with which a juror believes in any particular proposition (and consequently in the ultimate decision proposition). However, because of lack of empirical support for Bayesian calculations as a *description* of human judgment under uncertainty, we have adopted a set of simple assumptions that will allow us to perform calculations over a network of relationships and that we believe are closer to actual juror judgment processes. The main assumption is that at the time that an inferential conclusion is being considered as a potential story event, it is either regarded as certainly true (and therefore as data, e.g., P2), or as uncertain (and therefore as an hypothesis, e.g., P3, C), or as rejected and therefore sized to be a function of its role in the story and its relation to relevant world knowledge (we return to the subject of juror confidence in the section on certainty principles).



Figure 8.2. An abstract episode schema.

of the discourse. Stories appear to be organized into units that are often called *episodes* (Mandler, 1984; Pennington & Hastie, 1986; Rumelhart, 1977; Schank, 1975; Stein & Glenn, 1979; Trabasso & van den Broek, 1985). We show an abstract episode schema in Figure 8.2 that depicts a typical configuration of events in an episode; an episode should contain events which fulfill particular roles and are connected by certain types of causal relationships. In stories and in episodes, events considered to be *initiating events* cause characters to have psychological *responses* and to form *goals* that motivate subsequent *actions* which cause certain *consequences* and accompanying *states*. An example of an episode in the Johnson case is the following sequence: Johnson and Caldwell are in Gleason's bar. Caldwell's girlfriend, Sandra Lee, goes up to Johnson and asks him for a ride to the race track the next day (initiating events). Caldwell becomes angry (internal response), pulls his razor, and threatens Johnson (actions, note that goal is missing). Johnson backs off (consequence).

Stories may have further structure by virtue of the fact that each component of an episode may be an episode itself. For example, the entire episode above (characterized as Caldwell threatens Johnson) is the initiating event in one version of the Johnson story. In this version, the afternoon "threat" episode causes Johnson to be angry, and want to pay Caldwell back. Thus, a story may be thought of as a hierarchy of embedded episodes (Rumelhart, 1977; Trabasso & van den Broek, 1985). The highest level episode characterizes the most important features of "what happened." Components of the highest level episode are elaborated in terms of more detailed event sequences in which causal and intentional relations among subordinate story events are represented.

The structure of stories, according to our theory, plays an important role in the juror's comprehension and decision making processes. The story constructed by the juror will consist of some subset of the events and causal relationships referred to in the presentation of evidence, and additional events and causal relationships inferred by the juror. Some of these inferences may be suggested by the attorney and some may be constructed solely by the juror. Whatever their source, the inferences will serve to fill out the episode structure of the story. Thus, expectations about the kinds of information necessary to make a story tell the juror when important pieces of the explanation structure are missing and when inferences must be made. Knowledge about the structure of stories allows the juror to form an opinion concerning the completeness of the evidence, the extent to which a story has all its parts. Second, the structure of episodes in a story corresponds to the structure of our knowledge about human action sequences in the world. That is, story construction is a general comprehension strategy for understanding human action. Thus the juror can easily compare the structure that is being imposed on the evidence to already encoded prior knowledge. Finally, the hierarchical episode and causal structure of the story provides an "automatic" index of the importance of different pieces of evidence (Trabasso & Sperry, 1985). In the example above, the details of the embedded "threat" episode are subordinate in importance to the details of the top level episode that reveal what Johnson did in order to pay Caldwell back. However, this indexing of importance is something that emerges from the *structure* of the story.

*Certainty principles.* More than one story may be constructed by the juror. However one story will usually be accepted as the "best" story. And, the juror will have a level of confidence in that "best" story that may be quite high or quite low. The principles that determine acceptability of a story and the resulting level of confidence in the story, we call *certainty principles.* According to our theory, two certainty principles govern acceptance: *coverage* and *coherence.* An additional certainty principle, *uniqueness*, will contribute to confidence.

A story's coverage of the evidence refers to the extent to which the story accounts for evidence presented at trial. Our principle states that the greater the story's coverage, the more acceptable the story as an explanation of the evidence, and the greater confidence the juror will have in the story as an explanation, if accepted. An explanation that leaves a lot of evidence unaccounted for is likely to have a lower level of acceptability as the correct explanation. Poor coverage should lower the overall confidence in a story and consequently lower confidence in the decision.

A story's coherence and level of confidence also enter into its acceptability. However, coherence is a concept in our theory that has three com-

ponents: consistency, plausibility, and completeness. A story is consistent to the extent that it does not contain internal contradictions either with evidence believed to be true or with other parts of the explanation. A story is plausible to the extent that it corresponds to the decision maker's knowledge about what typically happens in the world and does not contradict that knowledge. A story is complete when the expected structure of the story "has all of its parts" (according to the rules of the episodic structure, see Figure 8.2 and discussion above). Missing information or lack of plausible inferences about one or more major components of the story structure will decrease confidence in the explanation. Thus, the coherence of the explanation reflects the consistency of the explanation with itself and with world knowledge, and the extent to which parts of the explanation can be inferred or assembled. These three ingredients of coherence (consistency, plausibility, and completeness) may be fulfilled to a greater or lesser degree and the values of the three components will combine to yield the overall level of coherence of a story.

Finally, if more than one story is judged to be coherent, then the stories will lack *uniqueness*, which contributes to confidence in a story and in a decision. If there are multiple coherent explanations for the available evidence, belief in any one of them over the others will be lessened (Einhorn & Hogarth, 1986; van Wallendael, 1989). If there is one coherent story, this story will be accepted as the explanation of the evidence and will be instrumental in reaching a decision. (These principles are elaborated and formalized in Pennington, Messamer, & Nicolich, 1991.)

*Summary.* Meaning is assigned to trial evidence through the incorporation of that evidence into one or more plausible accounts or stories describing "what happened" during events testified to at the trial. General knowledge about the structure of human purposive action sequences and of stories, characterized as an episode schema, serves to organize events according to the causal and intentional relations among them as perceived by the juror. Specific world knowledge about events similar to those in dispute will determine which particular interpretation is constructed or accepted. The level of acceptance will be determined by the coverage, coherence, and uniqueness of the "best" story.

#### Learning verdict definitions

The second processing stage in the juror's decision, according to the story model, involves the comprehension and learning of the decision alternatives, which in criminal trials are the definitions of the verdict alternatives (e.g., first-degree murder, second-degree murder, etc.). Most of the information for this processing stage is given to jurors at the end of the trial in the judge's substantive instructions on the law although jurors may also

#### FIRST DEGREE MURDER

IDENTITY:	Right Person
MENTAL STATE:	Intent to Kill Purpose Formed
CIRCUMSTANCES:	Insufficient Provocation Interval Between Resolution and Killing
ACTIONS:	Unlawful Killing Killing in Pursuance of Resolution

Figure 8.3. Example verdict category represented as a feature list.

have prior ideas concerning the meaning of the verdict categories (see Figure 8.1, bottom).

The verdict definition information in the judge's instructions is usually abstract and often couched in unfamiliar language: A crime is named and then abstract features are presented that define the crime. Features typically describe requirements of *identity, mental state, circumstances,* and *actions* that constitute the crime (Kaplan, 1978). For example, a judge's definition of first-degree murder presented as a feature list is shown in Figure 8.3.

We hypothesize that the juror's mental representation of this information also takes the form of a category label with a list of features. In all respects, this is a difficult one-trial learning task. If the juror has no prior knowledge of the legal categories, then learning of the abstract information is extremely difficult. In the case where prior knowledge is available, it is as likely to interfere with accurate understanding as to help, because jurors' prior exposures to concepts such as first degree murder, manslaughter, armed robbery, etc. are often (mis-) informed by television episodes and other media presentations.

#### Making a decision

The third processing stage that we hypothesize in the juror's decision making involves matching the accepted story with each of the verdict definitions. In cognitive processing terms, this is a classification process in which the best match between the accepted story's features and verdict category features is determined (see Figure 8.1, middle).

Because verdict categories are unfamiliar concepts, the classification of a story into an appropriate verdict category is likely to be a deliberate process. For example, a juror may have to reason about whether a circumstance in the story such as "pinned against a wall" constitutes a good match to a required circumstance, "unable to escape," for a verdict of not guilty by reason of self-defense. Although difficult, the classification process is aided by relatively direct relations between attributes of a verdict category (crime elements) and components of the episode schema (see Figure 8.4). The law has evolved so that the main attributes of the decision categories suggested by legal experts (Kaplan, 1978) – identity, mental state, circumstances, and actions – correspond closely to the central features of human action sequences represented as episodes: initiating events, goals, actions, and states. This is not a coincidence; rather, it is a reflection of the fact that both stories and crimes are culturally determined generic descriptions of human action sequences.

The story classification stage also involves the application of the judge's procedural instructions on the presumption of innocence and the standard of proof. If the best fit is above a threshold requirement, then the verdict category that matches the story is selected. If not all of the verdict attributes for a given verdict category are satisfied "beyond a reasonable doubt" by the events in the accepted story, then the juror should presume innocence and return a default verdict of not guilty. We are basing this hypothesis on the assumption that jurors either (A) construct a single "best" story, rejecting other directions as they go along or (B) construct multiple stories and pick the "best." In either case, we allow for the possibility that the best story is not good enough or does not have a good fit to any verdict option. Then a default verdict has to be available.

*Certainty principle.* A further assessment of confidence occurs in the story classification stage. An evaluation of *goodness-of-fit* between the story and the best-fitting verdict category is based on the extent to which the story includes instantiations of elements of the category. The more missing element matches between the components of the episode schema and the attributes of the verdict category (see Figure 8.4), the lower the juror's confidence in the verdict. As we proposed above, if the goodness-of-fit is not sufficient, then a default decision will be made.

#### Temporal relations between stages

The processing stages have been presented as though a story is constructed, then the verdicts are represented, and then a decision is reached. A fundamental claim of our theory is that the explanation structure is created a priori and *causes* the decision and is not a structure that is developed as a post hoc justification of the decision. This does not preclude a version of the theory in which there is cycling through the decision phases more than once; in such a case there could be an influence of the tentative decision (initial verdict classification) on an elaboration of the explanation. For example, story construction probably does not stop abruptly at the conclusion of the presentation of evidence. Previous research suggests that jurors' judgments involve much weighing and sifting of evidence as well



Figure 8.4. The main elements of a story (episode schema) map onto the defining attributes of a verdia (verdict category attributes).

as reflection on the meaning of the verdict categories, after the major courtroom body of the trial is concluded (Hastie, Penrod, & Pennington, 1983; Pennington, 1981). In addition, story meanings are not static structures. Although many causal and intentional inferences are made during the initial comprehension and encoding of events, causal information processing may not be completed during comprehension (e.g., Kintsch, 1974). Rather, causal reasoning begins then and continues with subsequent attributional inferences influenced by the initial comprehension process. Examining the extent to which processing stages interact is a long-term goal of our research.

#### Empirical evidence for the story model

The basic claim of the story model is that story construction enables critical interpretive processing and organization of the evidence so that evidence can be meaningfully evaluated against multiple verdict judgment dimensions. The story model provides a psychological account for the assignment of relevance to presented and inferred information.

Other writers have recognized the pervasive role of narrative structures in the comprehension of social events in our culture. Indeed, the story model of juror decision making accrues some support from the popularity of notions of story-telling and narrative structure in the writings of other scholars interested in legal practice: argumentation and persuasion at trial (Mauet, 1988; Moore, 1989; Schuetz & Snedaker, 1988; Twining, 1990); lawyer-client communication (Cunningham, 1989); plea bargaining (Mather, 1979; Maynard, 1988); litigant satisfaction with trial outcomes (O'Barr & Conley, 1985); comparative law (examples in Hamnett, 1977); and juror decision making (Bennett & Feldman, 1981; Holstein, 1985; Lopez, 1984).

The story model is distinctive among "narrative" approaches in the precision of its claims concerning the representational form of the evidence and the central role that is assigned to stories in subsequent decisions and confidence in those decisions. Uncertainty in the decision is based on assessments of the coverage, coherence, and uniqueness of the story and on the goodness-of-fit of the story with reference to the verdict categories. Detailed summaries of empirical studies of the claims of the story model are provided in other reports (Pennington, 1981, 1991; Pennington & Hastie, 1981b, 1986, 1988, 1992). In this section, we summarize the empirical support for the theory.

#### Interview study

Our initial research on the story model (Pennington, 1981; Pennington & Hastie, 1986) was designed to elicit data that would provide a snapshot of the juror's mental representations of evidence and of verdict categories at one point in time. Three questions were the focus of the empirical analysis: Do the mental representations of evidence show a story structure? Do the

mental representations of verdicts show a category structure? Are there systematic relationships between an individual juror's verdict decision and that same juror's evidence representation, verdict representation, or classification procedures? The first study was based on a correlational logic, using an interview to provide structural descriptions of mental representations and then determining whether or not the structures covaried systematically with verdict decisions. According to the story model, if story construction is a central determinant of verdict decisions, then we should find that variability in verdict decisions is correlated with variability in story structures and is not related to verdict category representations or classification procedures.

Twenty-six adult subjects were sampled from volunteers in the Massachusetts Superior Court jury pool and shown a realistic filmed reenactment of a murder trial (the *Commonwealth v. Johnson* case described earlier).

Subjects were instructed that an actual jury had decided the case and their task was to "be one of the jurors" and to try to reach a decision on the verdict. In the trial the jurors chose from among four verdict alternatives in reaching a decision. The defendant Frank Johnson could be judged to be not guilty, guilty of manslaughter, guilty of second-degree murder, or guilty of first-degree murder. A subsample of 16 of the 26 subjects was chosen for extensive analysis so that a range of verdicts was obtained (see Pennington, 1981; Pennington & Hastie, 1986 for details). The verdict distribution for the sixteen subjects was: 5 guilty of first-degree murder, 4 guilty of second-degree murder; 4 guilty of manslaughter; 3 not guilty (self-defense). The major source of data was a verbal protocol obtained in an interview with each experimental juror asking them to talk out loud while making a decision and to respond to questions about the evidence and about the judge's instructions to the jurors.

Verbal protocols from each juror were analyzed by coding all assertions about events and relationships between events that were claimed to have occurred or not to have occurred within the context of the events referred to in testimony. Many of these assertions referred to events and relationships actually mentioned in testimony and many referred to events and relationships inferred by the juror. This coding was transformed into a directed graph designating interrelations between events (Goodman & Hedetniemi, 1979).<sup>3</sup> A graph structure was created for each subject in which the nodes represented event codes and the links represented the asserted connections between events. This structure captured part of each subject's conceptual representation of the evidence as indicated by the protocol events mentioned and assertions regarding relations between events (Graesser &

<sup>&</sup>lt;sup>3</sup> A directed graph is a structure that includes points (nodes) and arcs (links) between points that have direction. In our application, the nodes will stand for events expressed as single states or actions, such as "Johnson was at the bar," and "Caldwell came over to Johnson's table." The links will stand for relationships between events. An example of one type of directed link would be an enabling causal relationship such as might exist between the two events above: Johnson's being at the bar "enabled" Caldwell to come over and talk to him.

Clark, 1985). To facilitate economy of presentation and to permit comparisons to other experiments, only the results from the two extreme verdict groups, first-degree murder and not guilty, will be summarized in the present report. Analyses of the data for subjects from all four verdict groups are in agreement with this summary (Pennington, 1981; Pennington & Hastie, 1986).

Before reviewing evidence that these graphs had a story structure, we should ask what is the range of plausible structures that we might expect? First, evidence could be stored in memory in an unembellished form as it was presented at trial, in a disconnected question and answer sequence, organized by witness and interconnected largely by referential coherence. This is plausible because we know that when judgments are made on-line<sup>4</sup> (Hastie & Park, 1986; Hastie & Pennington, 1989), memory for evidence is unrelated to the judgment. Second, the evidence could be conceptualized in terms of the structure of the legal argument (see Pennington & Hastie, 1981a, Figure 3, p. 256, for an example of such an evidence structure), as analyzed by legal scholars (e.g., Anderson & Twining, 1991; Wigmore, 1937) and other theorists (e.g., Schum & Martin, 1982). In this conception, evidence is structured in terms of arguments for and against guilt; or for and against a required element of guilt with respect to a particular charge. A third possibility is that the important evidence revolves around the characterizations of the defendant and victim. In this case structures emerging from protocols would show character sketches connected to verdicts. A final possibility is our theory, that the juror organizes the evidence into a story that emphasizes the causal and intentional relations among evidence items.

Our first major conclusion from the interview study was that the mental representations of evidence derived from the interview protocols showed story structures and not other plausible structures. There were several features of the conceptual graph structures that support our claim that these structures had story form and not one of the other plausible forms. First, 85% of all the events referred to in the protocols were causally linked. Thus, subjects were primarily making assertions like, "Johnson was angry so he decided to kill him," (anger initiates goal to kill) rather than assertions like, "Johnson was a violent man. That makes me think he intended to kill him." (The fact that Johnson was violent leads to an inference of intention to kill.) This is strong evidence that subjects were telling stories and not constructing arguments (see Olson, Duffy, & Mack, 1984 for empirical evidence on the psychological differences between narrative forms

<sup>&</sup>lt;sup>4</sup> Making a judgment "on-line" means incorporating the value of a piece of evidence into a judgment as soon as it is encountered. In a legal trial context, this means that a witness testifies, "Johnson was carrying a knife," and the juror immediately increases his or her belief in guilt. If the witness then says, "It was a fishing knife," the juror immediately decrements belief in guilt. Story construction is not an "on-line" decision strategy, but rather a "memory-based" strategy because evidence is organized, elaborated, and interpreted in memory before entering into a judgment.

and argument forms). Second, only 55% of the protocol references were to events that were actually included in testimony. The remaining 45% were references to inferred events – actions, mental states, and goals that "filled in" the stories in episode configurations. This argues strongly against the image of the juror as a "tape recorder" with a list of trial evidence in memory. Experimental jurors did make character inferences (5.4% of story content) but these were integrated into the story structures as reasons for certain behaviors. For example, an inference that Caldwell was a violent man might be given as a reason that Caldwell pulled a razor when provocation was slight or as a reason that Johnson was afraid. Finally, the conceptual graphs can be represented as hierarchies of embedded episodes when rules are applied to identify explicit goals linked to actions leading to final consequences (see Pennington, 1981). Examples of these structures are illustrated in Figure 8.5.

The second major conclusion from the interview study was that story structures differed systematically for jurors choosing different verdicts. In order to analyze this, a measure of shared features (Tversky, 1977; Tversky & Gati, 1978) was used to develop a central story for each verdict group. We call these *verdict stories;* for example, the central story for the jurors choosing first-degree murder is the first-degree murder verdict story. A network was assembled containing only those event codes and links shared in common by 80% of the members of the verdict group. An episode structure was imposed on the causal chains by applying rules to identify explicit goals linked to actions leading to the final consequence (Pennington, 1981). Verdict stories for first-degree murder and not guilty verdict groups are shown in Figures 8.5A and 8.5B.

The gist of the first-degree murder verdict story (Figure 8.5A) is that an argument and threat by Caldwell (the victim) so enraged Johnson (the defendant) that he decided to kill him. He got his knife, found Caldwell, got in a fight and stabbed him to death. In contrast, the gist of the not guilty story (Figure 8.5B) is that Caldwell started a fight with Johnson and pulled a razor on him. Johnson pulled a knife in order to protect himself and Caldwell ran onto the knife.

The episode structures of the two stories map neatly onto their respective verdict category attributes. For example, in the not guilty verdict story (Figure 8.5B), there are three episodes, two of which are embedded. The main episode is the fight, and the initiating events are all Caldwell's actions during the fight. The afternoon episode serves to fortify not guilty subjects' conclusions about Johnson's psychological state, leading first to a goal to show the knife and then to actively protect himself. The not guilty story shows the knife going into Caldwell as a consequence rather than as a goaldirected action. These features correspond to the verdict features of not guilty by reason of self-defense: under immediate attack, unable to escape, intent to defend, and reasonable retaliation. First-degree murder requires



Figure 8.5A (top). First-degree murder verdict story (central story for jurors choosing first-degree murder). Figure 8.5B (bottom). Not guilty verdict story (central story for jurors choosing not guilty by reason of self-defense).

premeditation, that is, a resolution formed to kill, an interval of time, and a killing in accordance with the resolution. The subjects' emphasis on the initiation of an intent-to-kill goal is expressed through the elaborated afternoon events (Figure 8.5A). Thus, being hit is not an initiating event, it is part of a sequence of acts that follow from behavior directed by a goal to kill.

Verdict representations were coded and compared across different verdict groups. The majority of references to verdicts during the talk aloud portion of the interview were to particular category features such as "premeditation" and "malice." When asked to tell what the verdicts meant, jurors provided lists of features although they were usually far from complete. Jurors also elaborated their category definitions in some cases by constructing ministories to illustrate. For example, "First-degree murder is premeditated. There is a plan. That would be if he had gone back to the bar looking for him."

There is considerable variation among jurors in the accuracy and completeness of their representations of the verdict category information. If these variations are systematically related to the jurors' verdict choices, then the hypothesis has to be retained that differences occurring in the verdict learning stage determine the verdict choice. For example, if jurors favoring first-degree murder verdicts were also jurors who did not remember the judge's instruction concerning premeditation, it would be plausible that the memory failure affected the verdict choice. To the contrary, analyses of the answers to questions about verdict category definitions showed that juror verdict choice was not related to memory for verdict-element relations. Other analyses (Pennington & Hastie, 1986) showed no content differences by verdict choice. Thus, variance in performance in the verdict learning stage of the juror decision does not appear to determine the juror's decision.

The interview study did not obtain information concerning the juror's notions of presumption of innocence or the beyond reasonable doubt standard. However other research sampling jurors from the same population of jurors and using the same stimulus trial (Hastie et al., 1983) did obtain direct ratings of subjects' estimated values for the beyond reasonable doubt threshold and the presumption of innocence principle. Analyses conducted on these ratings did not find significant differences among the subjects when they were classified by verdict choice. The implication for the present work is that variation in performance in the story classification stage is not associated with variations in verdict choice.

In this research, two key results were established that were necessary conditions for pursuit of the story model as a viable theory of decision making in the juror context. First, the evidence structures constructed by jurors had story structure (not other plausible structures) and verdict structures looked like feature lists. Second, jurors who chose different verdicts had constructed different stories. Thus, decisions covaried with story structures, but not with verdict representations or story classification processes.

The interview study served its purpose as our initial investigation and played a role in our theory building. Moreover, it provides a strong empirical foundation for the story model using a realistic stimulus trial and a range of adult citizen jurors. However, the interview methodology itself may have created a demand for stories as conversational forms or as justifications. Subsequent experiments addressed this issue and tested additional claims of the theory.

#### Predicting importance ratings and memory for evidence

A second empirical study (Pennington & Hastie, 1988) was conducted to test the conclusions of the interview study using conventional laboratory research methods with college student subjects. In this study subjects' responses to sentences presented in a recognition memory task were used to identify subjects' postdecision representations. The major motivation for this study was to test whether or not stories were constructed spontaneously in the course of the juror's performance. A second goal of the study was to test our claim that the position of an evidence item in its verdict story would predict the importance rating for that item according to verdict choice.

Stimulus materials were constructed based on the content analysis of interviews from the first study. This yielded a 119-sentence written version of the *Commonwealth v. Johnson* case. The written case was carefully constructed so that its constituent sentences included propositions from each verdict story that were not also present in any other verdict story. For example, the proposition "Johnson stabbed Caldwell" was a part of the first-degree murder verdict story, but not included in the not guilty verdict story.<sup>5</sup> In addition, sentences were prepared for the recognition test that were not presented in the stimulus case, but which were parts of the verdict stories identified in the interview study (that is they were frequently inferred by the jurors choosing a particular verdict). Thus, a recognition test had been presented as evidence, and with *new* (false) lures from each story that had not been presented as evidence (but were frequent inferences).<sup>6</sup>

<sup>&</sup>lt;sup>5</sup> This is an empirical observation. This particular statement could have, in principle, been in any of the verdict stories. We described earlier how verdict stories were determined empirically: The graph of events and links for each juror choosing the same verdict were compared. Those events and links in the stories of 80% of the jurors in the verdict group were retained as part of the verdict story. Some of the constituent events and links were evidence items and some were inferences shared by jurors in the verdict group. A particular verdict story would contain only part of the total body of evidence and only part of the total body of inferences that various jurors drew.

<sup>&</sup>lt;sup>6</sup> Examples of each type of recognition memory test item are as follows: Old items (presented as evidence) and empirically determined to be a part of:

#### 210 NANCY PENNINGTON AND REID HASTIE

Subjects "recognized," as having been presented as trial evidence, sentences from the story associated with their verdict with a higher probability than sentences from stories associated with opposing (rejected) verdicts, for both old and new items. That is, they were more likely to correctly recognize as evidence those evidence items in the verdict story corresponding to their own verdict choices, *and*, they were more likely to falsely recognize as evidence those inferences in the verdict story corresponding to their decisions. Subjects also rated the importance of trial evidence items and these ratings were strongly related to the causal role of the item in the story associated with a subject's verdict. These results corroborated the conclusions about story structure and story-verdict relationships from the initial study. Furthermore, they implied that story representations were constructed *spontaneously*, as part of the natural decision process, and not artificially elicited by the interview task used in the first study.

Even though we demonstrated with this experiment that causal explanations are constructed spontaneously in the course of decision making, we have still not demonstrated that the story constructed by the juror is a true mediator of the decision. It could still be the case that the juror makes a decision and then constructs a story as a post hoc justification. The next experiments address this question.

#### Predicting decisions and confidence in decisions

Two experiments were conducted to test our claim that stories cause decisions. We reasoned that if we could manipulate the ease of constructing a particular story and thereby influence the likelihood of the corresponding decision, then this would be strong evidence for our claim of causal mediation. In both studies, we manipulated the ease of constructing a particular story by varying the presentation order (but not the content) of the evidence.

In our third study (see Pennington & Hastie, 1988), using the abbreviated *Commonwealth v. Johnson* stimulus trial, we varied presentation order to influence the ease with which a prosecution (guilty of murder) or defense (not guilty by reason of self-defense) story could be constructed. Stories were considered easy to construct when the evidence was ordered in a temporal and causal sequence that matched the occurrence of the original events (*story order*). Stories were considered difficult to construct when the presentation order did not match the sequence of the original events. We based the nonstory order on the sequence of evidence as conveyed by

Not guilty verdict story: Johnson held his knife out in front of himself.

Murder verdict story: Johnson stabbed Caldwell in the chest.

New items (not presented as evidence) but empirically determined to be a frequent inference in:

Not guilty verdict story: Johnson was trying to protect himself.

Murder verdict story: Johnson was looking for Caldwell.

	Defense evidence (%)		
Prosecution evidence	Story order	Witness order	Means (%)
Story order	59	78	69
Witness order	31	63	47
Means	45	70	

 Table 8.1. Percentage of subjects choosing a verdict of guilty of murder by prosecution and defense order conditions

witnesses in the original trial (*witness order*). The logic of the experiment was summarized in our hypothesis that (manipulated) ease of story construction would influence verdict decisions; easy-to-construct stories would result in more decisions in favor of the corresponding verdicts.

One hundred and thirty college student mock jurors listened to a tape recording of a 100-item version of the *Commonwealth v. Johnson* evidence (50 prosecution statements and 50 defense statements), followed by a judge's charge to choose between a guilty of murder verdict and a not guilty verdict. The 50 prosecution statements, the first-degree murder story were presented either in story order or witness order. Similarly, the defense statements were presented in one of the two orders creating a four-cell factorial design. In all four order conditions the prosecution evidence preceded the defense evidence as per standard legal procedure. After listening to the tape recorded trial materials, the mock jurors completed a questionnaire indicating their verdict, confidence in the verdict, and their perceptions of the strengths of the prosecution and defense cases.

As predicted, mock jurors were likeliest to convict the defendant when the prosecution evidence was presented in story order and the defense evidence was presented in witness order (78% chose guilty) and they were least likely to convict when the prosecution evidence was in witness order and defense was in story order (31% chose guilty, see Table 8.1). Thus, story coherence, as determined by presentation order of evidence, affects verdict decisions in a dramatic way. (See Smith, 1988, for a replication of this study with different materials comparing story order with an organization of evidence by *legal issue* rather than with witness order.)

Analyses were conducted on the ratings of strength of the defense and prosecution cases and these ratings were influenced by presentation order, with story order evidence rated as stronger than witness order. Furthermore, the perceived strength of one side of the case depended on both the order of evidence for that side and for the other side of the case. This finding supports our claim that the *uniqueness* of the best-fitting story is one important basis for confidence in the decision. We also examined the verdict confidence ratings and found that, regardless of verdict chosen, jurors who heard *both* sides of the case in story order were more confident than jurors who heard one or neither side in story order. This result reinforces our conclusion that alternate story strength is important, although the empirical finding was not predicted.

It should be noted that this experiment was a laboratory experiment designed to test our hypotheses about the presence or absence of certain mental processes and their relationship to the decisions made. The study was not designed to estimate the *size* of order effects in real trials. In real trials, there are many devices that assist the juror in story construction: opening and closing statements, redundancy in presentation of information, a rich visual environment, and so forth. In this particular experiment, we stripped those enhancements away to reveal the effect of making a story very difficult to construct. In order to estimate the size of order effects in actual trials, this type of experiment would need to be repeated with more realistic stimulus materials.

A fourth experiment focused again on the effects of variations in evidence presentation order, allowing us to further examine the relationship between evidence organization, memory organization, recall memory, and judgments (see Pennington & Hastie, 1992). Using case materials developed by Devine and Ostrom (1985), evidence was presented either in *story order* or *legal issue order*. Two cases were used that varied in whether the preponderance of evidence favored guilt or innocence.

Materials were presented to 414 college student subjects in written form. The two evidence organizations resulted in different memory organizations of evidence, as measured by an analysis of clustering in free recall (using an "adjusted ratio of clustering," ARC, discussed in Ostrom, Pryor, & Simpson, 1981).<sup>7</sup> Subjects who heard the evidence organized by story showed high story clustering and low issue clustering in free recall; subjects who heard the evidence organized by legal issue showed low story clustering and high issue clustering. However, total amount recalled was not different for the two evidence organization conditions.

The results replicate and extend our previous results. When stories were easily constructed, and therefore represented more *coherently* in memory, more verdicts were chosen in the expected direction and subjects rated their confidence as higher. Moreover, this effect was obtained in the absence of effects on the overall amount of recall, ruling out the hypothesis that manipulating order merely manipulates memorability of the evidence.

We have now predicted and observed sizeable effects of story coherence on verdict choice and on confidence that are consistent with the story model. Furthermore, the effects appear in two very different sets of case materials.

<sup>&</sup>lt;sup>7</sup> This is computed by counting the number of times two statements in the same "story" are recalled together and expressing this number as a proportion of total recall, with both numerator and denominator adjusted for the expected number of the story items that would occur together by chance if recall were a random sample of items. Similarly, an ARC would be computed for items recalled together that referred to the same legal "issue."

Alternate algebraic models for juror judgment, derived from the Bayesian approach and primacy-recency principles of information integration, cannot account for the particular pattern of order effects we have obtained. In our next experiments, we utilized a new set of experimental materials and tested predictions from the story model, a Bayesian model, and a sequential updating model (Pennington & Hastie, 1981a).

#### Comparisons with other models

Many psychological and legal analyses of the juror's task postulate that the decision depends on the estimation and combination of probabilities. A variety of important legal concepts make reference to the probabilistic nature of evidence at trial. For example, relevant evidence is defined as evidence that has a tendency to make the existence of a fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence (Federal Rules of Evidence, Rule 401, 1976; Lempert, 1977).

There is no doubt that the juror's task involves uncertainty, as do all complex decision tasks of the kind we are considering. But, treating the task as a probability assessment task assumes that the uncertainty assessments behave in ways that are consistent with the rules of mathematical probability theory. Within the mathematical (Pascalian) probability system there is a prescription for coherent probability revision in the light of evidence (Bayes' rule). Under this prescription, probabilistic opinion revisions have three basic properties: the combining process is multiplicative, probabilities of alternate hypotheses must sum to one, and a hypothesis that is held at any time with probability of zero cannot be revived (Schum & Martin, 1980). It is generally known that the Bayesian system is an inadequate description of human behavior under most conditions (Fischhoff & Lichtenstein, 1978; Rapoport & Wallsten, 1972; Slovic & Lichtenstein, 1971).

An alternate probability model has been proposed by Cohen (1977), the inductive probability system, in which probabilities have only ordinal properties. Negation is not complementary, zero probabilities correspond to "no reason to believe" and can therefore be revived with further evidence, and the opinion revision process is not multiplicative. Schum and Martin (1980), in a recent test of the descriptive adequacy of both the Bayesian and inductive probability systems as theories for juror judgments conclude: "We can be fairly conclusive in saying that our subjects did not typically respond in accordance with the canons of probabilistic inference in either the Baconian or Pascalian probability systems" (p. 77). (See also Einhorn & Hogarth, 1985; Pennington & Hastie, 1988.)

In general, features of human uncertainty assessment found across many tasks are inconsistent with the rules of one or more of the traditional probability calculi. For example, the subjective probabilities of complementary hypotheses have been found not to sum to one (Edwards, 1962; Einhorn & Hogarth, 1985; Robinson & Hastie, 1985; Schum & Martin, 1980; van Wallendael, 1989; van Wallendael & Hastie, 1990); if certainty about one hypothesis increases, certainty about alternate hypotheses may remain constant, increase or decrease (Robinson & Hastie, 1985; Schum & Martin, 1980); hypotheses held with subjective certainty of zero are frequently "revived" (Schum & Martin, 1980); the subjective certainty attached to a conjunction of events is frequently overestimated relative to the optimal combination of the component uncertainties (Bar-Hillel, 1973; Goldsmith, 1978); indeed, the subjective certainty attached to a conjunction of events may be assessed to be greater than the certainty of one or more of the component events (Leddo, Abelson, & Gross, 1984; Tversky & Kahneman, 1983); subjective certainty assessments may be too high under conditions where there is a high similarity between the pattern of evidence and a known standard, or when there is high internal consistency of the evidence even though the evidence is known or thought to be unreliable (Saks & Kidd, 1980; Schum, DuCharme, & DePitts, 1973; Schum & Martin, 1982; Tversky & Kahneman, 1974).

The alternative to modeling juror inference as a probabilistic opinion revision process is to consider that the weight of the evidence "accumulates" in some other manner. In this regard, the additive models (Information Integration Theory averaging rule and sequential weighting, reviewed by Pennington & Hastie, 1981a) are more consistent with the anomalies listed above than are the probability formulations.

In our final two experiments on the story construction process, we examined more closely the impact of story completeness on subjects' beliefs in the guilt of the defendant and its effect on evidence evaluation when subjects were asked to respond to the evidence at different levels of aggregation (see Pennington & Hastie, 1992). We expected that more complete stories would produce more verdicts in the direction of the completed story. We also expected that there would be a greater effect of mediating story structures when evidence was evaluated globally at the end of all of the evidence compared to judgments rendered after each item of evidence was presented. We also compared these two decision modes, subjects' *global* judgments<sup>8</sup> (the normal decision mode for legal judgments) and their cumulative *item-by-item* judgments,<sup>9</sup> with Bayesian, sequential updating, and story models of aggregation.<sup>10</sup>

- <sup>8</sup> A global judgment will refer to the condition in which subjects read through the entire body of evidence and made a single evaluation of the likelihood of guilt at the end of that reading.
- <sup>9</sup> Cumulative *item-by-item* judgments refer to the judgment condition in which subjects were asked to read a single block of evidence and then make a judgment of the likelihood of guilt, then read the next block, then make a new judgment (based on all evidence up to that point).
- <sup>10</sup> We compared these models to subjects' actual judgments by including a third *local* judgment condition in which subjects were asked to rate the probative value of each evidence block independently. We then applied the three model combination rules to these local judgments and compared the model aggregation result to the subjects' actual global and item-by-item judgments. The Bayes combination rule is well known; each evidence block

We directly varied the ease of constructing particular stories by providing or withholding evidence that was of specific relevance to one possible story or another. These "evidence supplements" were designed to instantiate a component of either the defense or the prosecution story by strengthening causal links between certain pieces of evidence and/or weakening others. In this way, we expected to alter the interpretation of the evidence, thus leading to different decisions. The case materials and methods in these two experiments were based on work by Schum and Martin (1982); three evidence conditions were created for two of their cases involving an embezzlement and a burglary: a convict supplements version, an acquit supplements version, and the original materials from Schum and Martin (basic version). For the first experiment, following methods laid down by Schum & Martin (1982), we had subjects respond to the case materials at three levels: a global assessment of the entire collection of evidence; local assessments of each block of evidence (essentially each witness's testimony); and an item-by-item evaluation where the subject responded after each block of evidence indicating his or her current cumulative judgment.

Because the supplements tied evidence together into a more (or less) coherent story, we expected that their effect would be greater when considered in the context of all the evidence (global judgment) than when their impact was incorporated into the judgment as the evidence was heard (item-by-item judgment). This prediction was motivated by the assumption that when subjects are asked to make a single global judgment after reading the entire body of evidence, they are able to integrate evidence into a unitary summary structure before evaluation, that is, their judgment strategy will be "memory-based" (Hastie & Park, 1986; Hastie & Pennington, 1989). However, when subjects are asked to make a cumulative judgment after each evidence block, the subject is focussed on the adjustment or change in evaluation. This is likely to invoke an "on-line" strategy whereby subjects anchor on the current opinion and adjust for the new evidence confronting them (Einhorn & Hogarth, 1985; Hastie & Park, 1986; Hastie & Pennington, 1989; Lopes, 1982; Robinson & Hastie, 1985). We also expected that neither the global nor the item-by-item judgments would be well fit by a Bayesian aggregation of the local evidence evaluations; that global judgments were more likely to have involved story construction; and that the item-by-item judgments would be better described by an anchor-and-adjust process.

Over the two experiments (see Pennington & Hastie, 1992), our prediction that the addition of story supplements would cause subjects to render stronger evaluations of evidence in the story direction was supported,

was considered to be independent (see formal analyses of these stimuli by Schum & Martin, 1980, 1982). The sequential updating model was an anchor-and-adjust model in which the current judgment (which is a summary of all previous judgments) was weighted .45 and the current evidence block was weighted .55. This model predicts large recency effects. The story model combination rule used equal weighting of evidence, that is, the probative evaluation of the item was its effective weight.

with the convict version of the cases eliciting greater odds in favor of guilt than the basic version and the basic version eliciting greater odds in favor of guilt than the innocent version.

Next, we tested our assumptions about the strategies subjects were using at different levels of aggregation. As predicted, the Bayesian model did not fare well as a description of subjects' global or item-by-item ratings in the experiment (also noted by Schum & Martin, 1982). First, consistent with an hypothesis of "conservatism," neither the final item-by-item nor the global ratings show the degree of influence of the evidence supplement manipulation that appears in the Bayesian calculation based on local evidence block ratings. Bayesian aggregates of the local judgments were about 10 times stronger than the global evaluations and about 15 times stronger than the item-by-item assessments. Thus the subject aggregates (global and item-by-item), consistent with previous research (Edwards, 1968; Schum & Martin, 1982), are extremely conservative with respect to a Bayesian aggregation rule. Second, several specific qualitative characteristics of the item-by-item ratings, primarily in the form of non-complementary adjustments, contradict implications of the Bayesian rule (see also Pennington & Hastie, 1988; Robinson & Hastie, 1985; Schum & Martin, 1982). Third, direct comparisons of goodness-of-fit of a Bayesian updating model and an algebraic anchor-and-adjust model (Einhorn & Hogarth, 1985; Lopes, 1982), applied to the item-by-item ratings, clearly favored the anchor-andadjust model. The mean difference between item-by-item ratings and the anchor-and-adjust model over evidence blocks is not reliably different from zero for either stimulus case. In contrast, the best fitting algebraic description of the global ratings was neither the Bayesian nor the anchor-andadjust model (differences between global ratings and anchor-and-adjust predictions were reliably different from zero). A configuration of weights consistent with the story model (see Pennington & Hastie, 1992) provided the best-fitting model for the global judgments.

We also predicted that story supplements would have greater impact on global judgments than on item-by-item judgments. This was supported by the fact that subjects' global assessments are stronger in force than the item-by-item final evaluation by a factor of about 1.5, and the predicted interaction between evidence supplement treatments (convict versus acquit) and response modes (item-by-item versus global) on final judgments of guilt was obtained.

In sum, the essential results of the two studies were consistent with predictions from the story model and projections from closely related research. The Bayesian model did not provide an adequate description of human performance on either the final ratings of the global judgment task or the ultimate rating of the item-by-item response sequence. Nor did the Bayesian approach provide an acceptable description of item-by-item ratings across the course of evidence presentation. An anchor-and-adjust algebraic updating model did provide a satisfactory fit to the sequence of item-by-item judgments. The final item-by-item judgment was less polarized (as a function of the presence of acquit or convict evidence supplements) in all conditions than the single global rating, as predicted from our hypothesis that anchor-and-adjust described the item-by-item judgment process, but story construction best described the global judgment.

*Summary.* The first study used an extensive interview to establish that intervening narrative structures were created by jurors in a realistic mockjuror study; that these structures took the form of a story; that jurors who agreed on the verdict decision shared a common story; and that other traces of the decision process (e.g., estimates of standard of proof, knowledge of the verdict definitions) did not covary systematically with the decision. The second study using a recognition memory task reinforced the first study's conclusions and added the finding that the story structures were created spontaneously, without the demands of communication with the experimenter in the interview situation.

The next two empirical studies provided substantial evidence that the storylike evidence summary is a key causal mediator of the verdict decision. In both studies variations in the order of presentation of a fixed set of evidence had clear effects on verdict decisions. Furthermore, the order manipulations were selected to facilitate or impede construction of convictionprone or acquittal-prone stories yielding successful predictions of verdicts from evidence order via the story model. The overall pattern of verdict decisions, confidence ratings, and other collateral judgments also supported our hypotheses that completeness, coherence, and uniqueness of the bestfitting story would predict confidence in the correctness of the verdict.

The final empirical studies provided some comparisons of the story model to two traditional computation-oriented models, a Bayesian updating model and an algebraic anchor-and-adjust model. At the most general level, we hypothesized that the Bayesian formulation would not provide a satisfactory account of any of the human judgments; that the story model would describe global judgments based on all of the evidence; and that the anchor-and-adjust model would describe the sequence of judgments when subjects were prompted for cumulative ratings after each witness's testimony had been summarized. The general hypothesis and subsidiary hypotheses derived from the story model and the anchor-and-adjust model were confirmed.

#### Conclusions

We have proposed a theory, the story model, that describes jurors' cognitive processing during the trial that results in a tentative predeliberation decision. The essence of this theory is that the construction of a causal model of events, a story, is central in understanding the evidence and its implications. Other processes, such as understanding the judge's instructions and matching the story with a decision option are necessary to turn the juror's understanding of the evidence into a decision.

#### 218 NANCY PENNINGTON AND REID HASTIE

We have also presented substantial empirical evidence supporting our assertions about the cognitive processing of jurors. Our current research addresses aspects of the theory for which we have not provided compelling empirical support. For example, we are currently investigating issues of generalizability - the extent to which the cognitive processes and mental structures proposed in the story model apply to a large range of legal cases and issues related to the generalizability of our theoretical principles to actual trial settings. A second focus of our current research concerns the time course of comprehension during decision making. By understanding which inferences are made when during the comprehension of evidence, the judge's instructions and during subsequent decision making, we will know whether jurors construct single or multiple stories; what factors influence the point at which alternative stories cease to be considered; and the extent to which processing stages interact to produce a decision. Another very important direction for development of the story model involves elaborating and formalizing the principles that we suggest determine confidence in decisions: coverage, coherence (completeness, consistency, and plausibility), uniqueness, and goodness-of-fit, and to formalize these principles in order to understand how confidence in a decision can result from a computation across semantic features of a mental representation of evidence.

In sum, we have conducted a long series of investigations on the story model and believe that it has been established as a major candidate to explain and predict juror decision making in criminal trials. Of course it is clear that there are many areas for further theoretical development and empirical research. Perhaps the most satisfying characteristic of the story model approach, for us as cognitive experimental psychologists, is the extent to which it connects important naturally occurring decision making phenomena to accounts from the mainstream of modern information processing theories of the mind.

#### References

- Anderson, T. J., & Twining, W. (1991). Analysis of evidence. Boston, MA: Little, Brown.
- Bar-Hillel, M. (1973). On the subjective probability of compound events. Organizational Behavior and Human Performance, 9, 396-406.
- Bennett, W. L., & Feldman, M. S. (1981). Reconstructing reality in the courtroom: Justice and judgment in American culture. New Brunswick, NJ: Rutgers University Press.
- Cohen, L. J. (1977). The probable and the provable. New York: Oxford University Press.
- Collins, A. M. (1978). Fragments of a theory of human plausible reasoning. In D. Waltz (Ed.), Proceedings of the Conference on Theoretical Issues in Natural Language Processing II (pp. 194-201). Urbana: University of Illinois Press.
- Collins, A. M., Brown, J. S., & Larkin, K. (1980). Inference in text understanding. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading* comprehension (pp. 385-407). Hillsdale, NJ: Erlbaum.

- Collins, A., & Michalski, R. (1989). The logic of plausible reasoning: A core theory. *Cognitive Science*, 13, 1–49.
- Crothers, E. J. (1979). Paragraph structure inference. Norwood, NJ: Ablex.
- Cunningham, C. D. (1989). A tale of two clients: Thinking about law as language. Michigan Law Review, 87, 2459-2494.
- Devine, P. G., & Ostrom, T. M. (1985). Cognitive mediation of inconsistency discounting. Journal of Personality and Social Psychology, 49, 5-21.
- Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review*, 69, 109-135.
  - (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York: Wiley.
- Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review*, 92, 433-461.
- Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. Psychological Bulletin, 99, 3–19.
- Ellsworth, P. C. (1988). Unpleasant facts: The Supreme Court's response to empirical research on capital punishment. In K. C. Haas & J. A. Inciardi (Eds.), *Challenging capital punishment: Legal and social science approaches* (pp. 177-211). Newbury Park, CA: Sage.
- Fischhoff, B., & Lichtenstein, S. (1978). Don't attribute this to Reverend Bayes. *Psy*chological Bulletin, 85, 239-243.
- Goldsmith, R. W. (1978). Assessing probabilities of compound events in a judicial context. *Scandinavian Journal of Psychology*, *19*, 103–110.
- Goodman, S. E., & Hedetniemi, S. T. (1979). A descriptive introduction to graph theory and some of its applications. In W. A. Sedelow, Jr. & S. Y. Sedelow (Eds.), *Computers in language research: Formal methods* (pp. 19–95). The Hague: Mouton.
- Graesser, A. C., & Clark, L. F. (1985). Structures and procedures of implicit knowledge. Norwood, NJ: Ablex.
- Hamnett, I. (Ed.) (1977). Social anthropology and law. London: Academic Press.
- Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, 93, 258–268.
- Hastie, R., & Pennington, N. (1989). Notes on the distinction between memorybased versus on-line judgments. In J. N. Bassili (Ed.), On-line cognition in person perception (pp. 1–17). Hillsdale, NJ: Erlbaum.
- Hastie, R., Penrod, S. D., & Pennington, N. (1983). *Inside the jury*. Cambridge, MA: Harvard University Press.
- Holstein, J. A. (1985). Jurors' interpretations and jury decision making. Law and Human Behavior, 9, 83-100.
- Kaplan, J. (1978). Criminal justice: Introductory cases and materials (2nd edition). Minola, NY: Foundation Press.
- Kintsch, W. (1974). The representation of meaning in memory. Hillsdale: NJ: Erlbaum.
- Leddo, J., Abelson, R. P., & Gross, P. H. (1984). Conjunctive explanations: When two reasons are better than one. *Journal of Personality and Social Psychology*, 47, 933–943.
- Lempert, R. O. (1977). Modeling relevance. Michigan Law Review, 75, 1021-1057.
- Lopes, L. L. (1982). Toward a procedural theory of judgment (Technical Rep. No. 17). Madison: University of Wisconsin, Wisconsin Human Information Processing Program.

#### 220 NANCY PENNINGTON AND REID HASTIE

- Lopez, G. P. (1984). Lay lawyering. UCLA Law Review, 32, 1-60.
- Mandler, J. M. (1984). Stories, scripts and scenes: Aspects of schema theory. Hillsdale, NJ: Erlbaum.
- Mather, L. M. (1979). Plea bargaining or trial? The process of criminal-case disposition. Lexington, MA: Lexington Books.
- Mauet, T. A. (1988). Fundamentals of trial techniques (2nd ed). Boston, MA: Little, Brown.
- Maynard, D. W. (1988). Narratives and narrative structure in plea bargaining. Law and Society Review, 22, 449-481.
- Moore, A. J. (1989). Trial by schema: Cognitive filters in the courtroom. UCLA Law Review, 37, 273-341.
- O'Barr, W. M., & Conley, J. M. (1985). Litigant satisfaction versus legal adequacy in small claims court narratives. *Law and Society Review*, 19, 661–701.
- Olson, G. M., Duffy, S. A., & Mack, R. L. (1984). Thinking-out-loud as a method for studying real-time comprehension processes. In D. E. Kieras and M. A. Just (Eds.), New methods in reading comprehension research (pp. 253-286). Hillsdale, NJ: Erlbaum.
- Ostrom, T. M., Pryor, J. B., & Simpson, D. D. (1981). The organization of social information. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), Social cognition: The Ontario Symposium on Personality and Social Psychology (Vol. 1, pp. 1-38). Hillsdale, NJ: Erlbaum.
- Pennington, N. (1981). Causal reasoning and decision making: The case of juror decisions. Unpublished doctoral dissertation, Harvard University.
  - (1991). A cognitive model of explanation-based decision making. Manuscript submitted for publication.
- Pennington, N., & Hastie, R. (1980, June). Representation and inference in juror reasoning: Two illustrative analyses. Paper presented at the annual meeting of the Cognitive Science Society, New Haven, CT.
- Pennington, N., & Hastie, R. (1981a). Juror decision-making models: The generalization gap. Psychological Bulletin, 89, 246-287.
  - (1981b, August). Juror decision making: Story structure and verdict choice. Paper presented at the meeting of the American Psychological Association, Los Angeles, CA.
  - (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, 51, 242–258.
  - (1988). Explanation-based decision making: Effects of memory structure on judgment. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 521-533.
  - (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology*, 62, 189–206.
- Pennington, N., Messamer, P. J., & Nicolich, R. (1991). Explanatory coherence in legal decision making. Manuscript in preparation.
- Rapoport, A., & Wallsten, T. S. (1972). Individual decision behavior. Annual Review of Psychology, 23, 131–176.
- Robinson, L. B., & Hastie, R. (1985). Revisions of beliefs when a hypothesis is eliminated from consideration. *Journal of Experimental Psychology: Human Perception* and Performance, 11, 443–456.
- Rumelhart, D. E. (1977). Understanding and summarizing brief stories. In D. La-Berge & S. J. Samuels (Eds.), Basic Processes in Reading: Perception and Comprehension. Hillsdale, NJ: Erlbaum.

- Saks, M. J., & Kidd, R. F. (1980). Human information processing and adjudication: Trial by heuristics. Law and Society Review, 15, 123–160.
- Schank, R. C. (1975). The structure of episodes in memory. In D. G. Bobrow & A. Collins (Eds.), Representation and understanding: Studies in cognitive science (pp. 237–272). New York: Academic Press.
- Schuetz, J. E., & Snedaker, K. H. (1988). Communication and litigation: Case studies of famous trials. Carbondale: Southern Illinois University Press.
- Schum, D. A., DuCharme, W. M., & DePitts, K. E. (1973). Research on human multistage probabilistic inferences processes. Organizational Behavior and Human Performance, 10, 318-348.
- Schum, D. A., & Martin, A. W. (1980). Probabilistic opinion revision on the basis of evidence at trial: A Baconian or a Pascalian process (Research Rep. 80-02). Houston, TX: Rice University, Department of Psychology.
  - (1982). Formal and empirical research on cascaded inference in jurisprudence. Law and Society Review, 17, 105-151.
- Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 6, 649-744.
- Smith, V. L. (1988). The psychological and legal implications of pre-trial instructions in the law. (Doctoral dissertation, Stanford University, 1987). *Dissertation Abstracts International*, 48, 3451B.
- Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing* (Vol. 2, pp. 83-107). Norwood, NJ: Ablex.
- Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. Journal of Memory and Language, 24, 595-611.
- Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, 24, 612–630.
- Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352.
- Tversky, A., & Gati, I. (1978). Studies of similarity, In E. Rosch & B. B. Lloyd (Eds.), Cognition and categorization (pp. 79–98). Hillsdale, NJ: Erlbaum.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124–1131.
- (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293-315.
- Twining, W. (1990). Lawyer's stories. In W. Twining, *Rethinking evidence* (pp. 219-261). Oxford: Basil Blackwell.
- van Wallendael, L. R. (1989). The quest for limits on noncomplementarity in opinion revision. Organizational Behavior and Human Decision Processes, 43, 385–405.
- van Wallendael, L. R., & Hastie, R. (1990). Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing. *Memory and Cognition*, 18, 240-250.
- Wigmore, J. 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.