Cognition, 49 (1993) 123–163

0010-0277/93/$06.00

Reasoning in explanation-based decision making

Nancy Pennington*, Reid Hastie
Psychology Department, Campus Box 345, University of Colorado, Boulder, CO 80309, USA

Abstract

A general theory of explanation-based decision making is outlined and the multiple roles of inference processes in the theory are indicated. A typology of formal and informal inference forms, originally proposed by Collins (1978a, 1978b), is introduced as an appropriate framework to represent inferences that occur in the overarching explanation-based process. Results from the analysis of verbal reports of decision processes are presented to demonstrate the centrality and systematic character of reasoning in a representative legal decision-making task.

Introduction

Many important decisions are made under conditions where a large base of implication-rich, conditionally dependent pieces of evidence must be evaluated as a preliminary to choosing a course of action. We propose that a model of *explanation-based* decision making describes behavior under these conditions (Pennington, 1981; Pennington & Hastie, 1981, 1986, 1988, 1991, 1992, 1993). According to the explanation-based model, decision makers begin their decision process by constructing a causal model to explain the available facts. The decision maker is also engaged in a separate activity to learn, create, or discover a set of alternatives from which an action will be chosen. A decision is made when the causal model of the evidence is successfully matched to an alternative in the choice set. The three processing stages in the general explanation-based decision model are shown in Fig. 1.

Distinctive assumptions in our explanation-based approach to decision making are the hypotheses that decision makers *reason about* the evidence in order to

*Corresponding author.

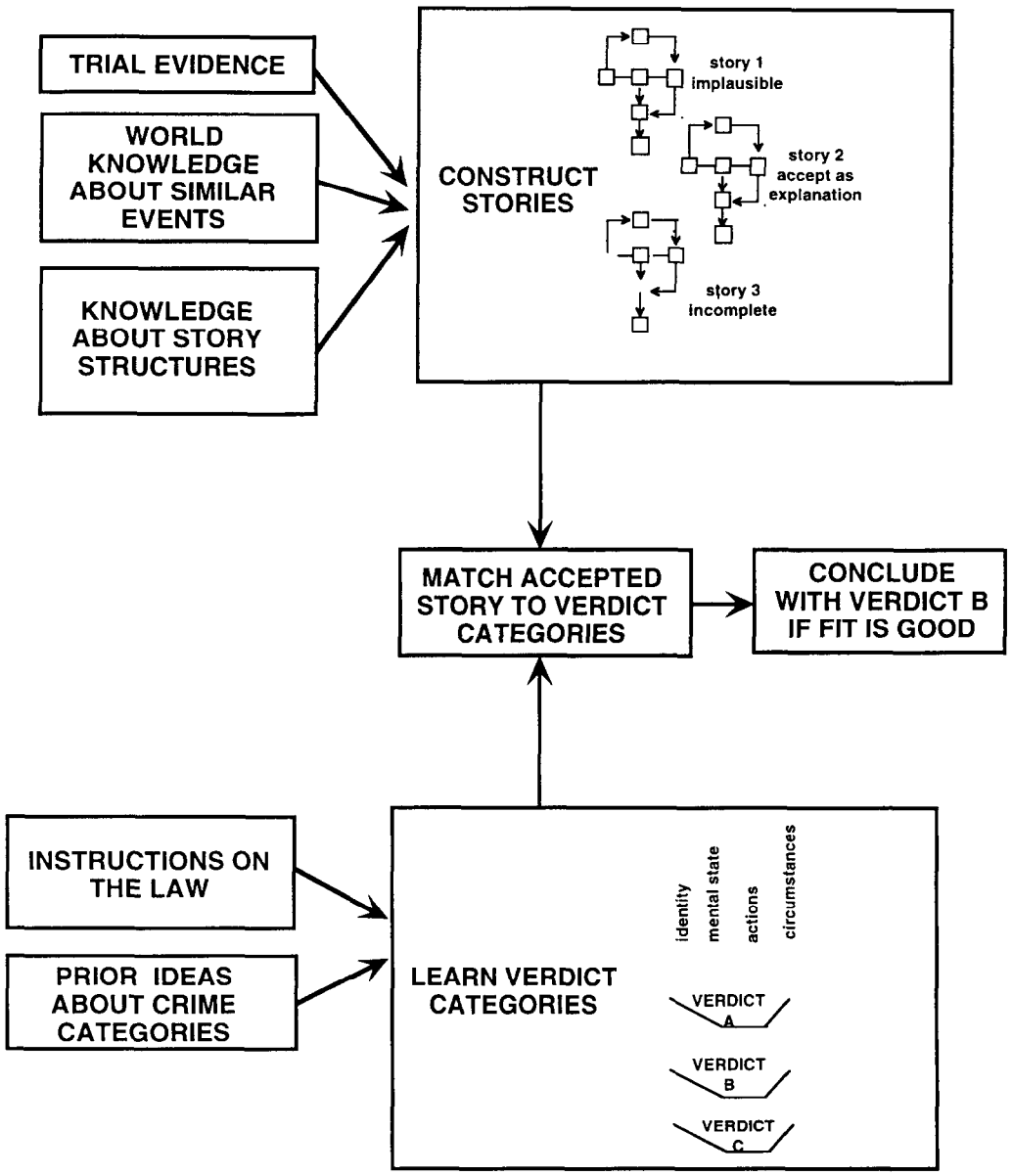

Figure 1. *Overview of the processing stages of the explanation-based model.*

construct an intermediate summary representation of the evidence, that this intermediate representation is an *interpretation* of the original evidence, and that this representation, rather than the original "raw" evidence, is the basis of the final decision. Interposition of this organized representation facilitates evidence comprehension, directs inferencing, enables the decision maker to reach a

decision, and contributes substantially to the confidence assigned to the accuracy or success of the decision. Our approach to decision processes parallels work demonstrating the role of explanation in categorization behavior (Murphy & Medin, 1985; Rips, 1989a), category learning (e.g., Schank, Collins, & Hunter, 1986), planning (Wilensky, 1983), and learning by generalization from examples (e.g., Lewis, 1988).

According to our theory of explanation-based decision making, the structure of the causal model constructed to explain the evidence will depend on the decision domain. Domain-specific causal rules and knowledge structures will underlie an internist's causal model of a patient's medical condition and its precedents (Pople, 1982), an engineer's mental model of an electrical circuit (de Kleer & Brown, 1983), an epidemiologist's model of the effect of radon contamination on health (Bostrom, Fischhoff, & Morgan, 1992), a merchant's image of the economic factors in a resort town (Hogarth, Michaud, & Mery, 1980), or a diplomat's causal map of the political forces in the Middle East (Axelrod, 1976; and see Klein, Orasanu, Calderwood & Zsambok, 1993, for additional examples). In the case of juror decision making we have established that a juror uses narrative story structures to organize and interpret evidence in criminal trials (Pennington & Hastie, 1986, 1988). In the present paper we will go beyond discussions of the *structure* of the model generated from the evidence to examine some of the *reasoning processes* that are involved in the creation and utilization of that mental representation.

The juror's decision task

The juror's decision task is a prototype of the tasks to which the explanation-based model should apply: First, a massive "database" of evidence is encountered at trial, frequently requiring several days to present. Second, the evidence comes in a scrambled order in which several witnesses and exhibits convey pieces of a historical puzzle in a jumbled temporal sequence. Third, the evidence is piecemeal and there are many gaps in its depiction of the historical events that are the focus of reconstruction: event descriptions are incomplete, usually some critical events were not observed by the available witnesses, and information about personal reactions and motivations is not presented (often because of the rules of evidence). Finally, subparts of the evidence (e.g., individual sentences or statements) are interdependent in their probative implications for the verdict. The meaning of one statement cannot be assessed in isolation because it depends on the meanings of related statements.

The general explanation-based model accounts for decision making by jurors through the following three component processes corresponding to those shown in Fig. 1: (a) constructing an explanation that accounts for the evidence in the form

of a "story"; (b) representation of the decision alternatives by learning verdict category attributes; and (c) reaching a decision through the classification of the story into the best-fitting verdict category. 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 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.

Constructing an explanation

Our empirical research has focused on the claim that the juror's explanation of legal evidence takes the form of a "story" in which causal and intentional relations among events are prominent (see also Bennett & Feldman, 1981; Hutchins, 1980; Pennington, 1981; Pennington & Hastie, 1986, 1988, 1992). Furthermore, we have shown that, under the conditions that hold in a typical criminal jury trial, jurors spontaneously construct story structures (and not other plausible structures); jurors who choose different verdicts construct different stories; and the story summary is a *cause* of the juror's decision.

According to the theory, the story will be constructed from three types of knowledge: (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. 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.

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's or reader's "mental representations" 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). 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 a 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.

Learning the choice set

The decision maker's second major task is to learn or to create a set of potential solutions or action alternatives that constitute the choice set. In criminal trials the legal information for this processing stage is given to jurors at the end of the trial in the judge's instructions on the law and the verdict alternatives available to the juror. These instructions provide only a sketchy outline of the decision categories and jurors may also have prior ideas concerning the meaning of the verdict alternatives. 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).

Again, constructive inference processes are rampant and it is common for prior conceptions of the verdicts (e.g., from news media and fictional accounts of trials) to intrude into the verdict representations (see Smith, 1991, for additional empirical results on verdict representation). But, many gaps and errors remain in the jurors' operative conceptualizations of the law (cf. Elwork, Sales, & Alfini, 1977; Hastie, Penrod, & Pennington, 1983).

Matching the story to the verdicts

The final stage in the global decision process involves matching decision alternatives to the summary evidence representation to find the most successful pairing. Because verdict categories are unfamiliar concepts, the classification of a

story into an appropriate verdict category is likely to be a deliberate inferential process. For example, a juror may have to decide 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. In this example, these inferences would depend on knowledge from the trial evidence, from the judge's instructions, and from the juror's background knowledge of human motivations (was the person "trying" to escape?), mental processes (was the person incapacitated?), and the physical world (was it physically possible for the person to escape?).

The story classification stage also involves the application of the judge's procedural instructions on the presumption of innocence and the standard of proof. That is, if not all of the verdict attributes for a given verdict category are satisfied "beyond a reasonable doubt", by events in the accepted story, then the juror should presume innocence and return a default verdict of not guilty.

Confidence in decisions

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 (see Pennington, Messamer, & Nicolich, 1991, for elaboration and formalization of these principles; see Collins, Brown, & Larkin, 1980, for a similar set of principles proposed to determine global confidence in one alternative interpretation of an ambiguous text).

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* also enters into its acceptability, and level of confidence given that the story is accepted. However, coherence is a concept in our theory that has three components: *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 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. Combination of these ingredients, however, is not strictly additive (Pennington, Messamer, & Nicolich, 1991). For example, completeness interacts with plausibility. If a story is plausible, then the completeness increases confidence. However, if a story is implausible, completeness does not increase confidence (it might be thought that completeness of an implausible story would actually decrease confidence but this is not the case; see Pennington, Messamer, & Nicolich, 1991).

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 (Baltzer & Pennington, 1993; 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 decisions.

Reasoning to construct representations

It is a trite but profound insight that "perception is only half of perception". The inevitable gaps in our sensory connections with the outside world are filled with information that is inferred. When we attempt to perceive the world second- or third-hand, as a juror or judge must, the gaps are larger and the role of inference is even more dominant. If we listen to jurors thinking aloud on their way to a decision, retrospectively reflecting on their verdicts, or discussing their decisions in a social context such as the jury room, we observe a flood of inferences; most of them directed at constructing a summary of "what happened", but many also drawn in support of a deliberate classification of the story into a verdict category. In this paper, we will focus on the reasoning to construct a story representation that accounts for as much of the trial evidence as possible.

We have hypothesized that jurors *impose* a narrative organization on evidence. By this, 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 (Crothers, 1979; Collins, Brown & Larkin, 1980; Kintsch, 1974, 1988).[1] This is especially true in the context of legal trials in which characteristics of the trial evidence make comprehension unwieldy. 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.

Our concept of reasoning to construct story representations is shown in Fig. 2. According to the theory, 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. For example, an evidence item, "Caldwell was in Gleason's Bar", is direct testimony, is not a matter of dispute, and it appears in all jurors' individual stories. In Fig. 2, this is shown as a piece of evidence (e1) appearing directly in the story. 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, 1978a, 1978b; Collins & Michalski, 1989). For example, 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. (from world knowledge)
P2. Caldwell was big. (from evidence)
P3. Caldwell was known to be a troublemaker. (previous inference)

C. Johnson was afraid. (inferential conclusion)

[1]This is a dominant view in cognitive psychology today. To illustrate, 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 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.

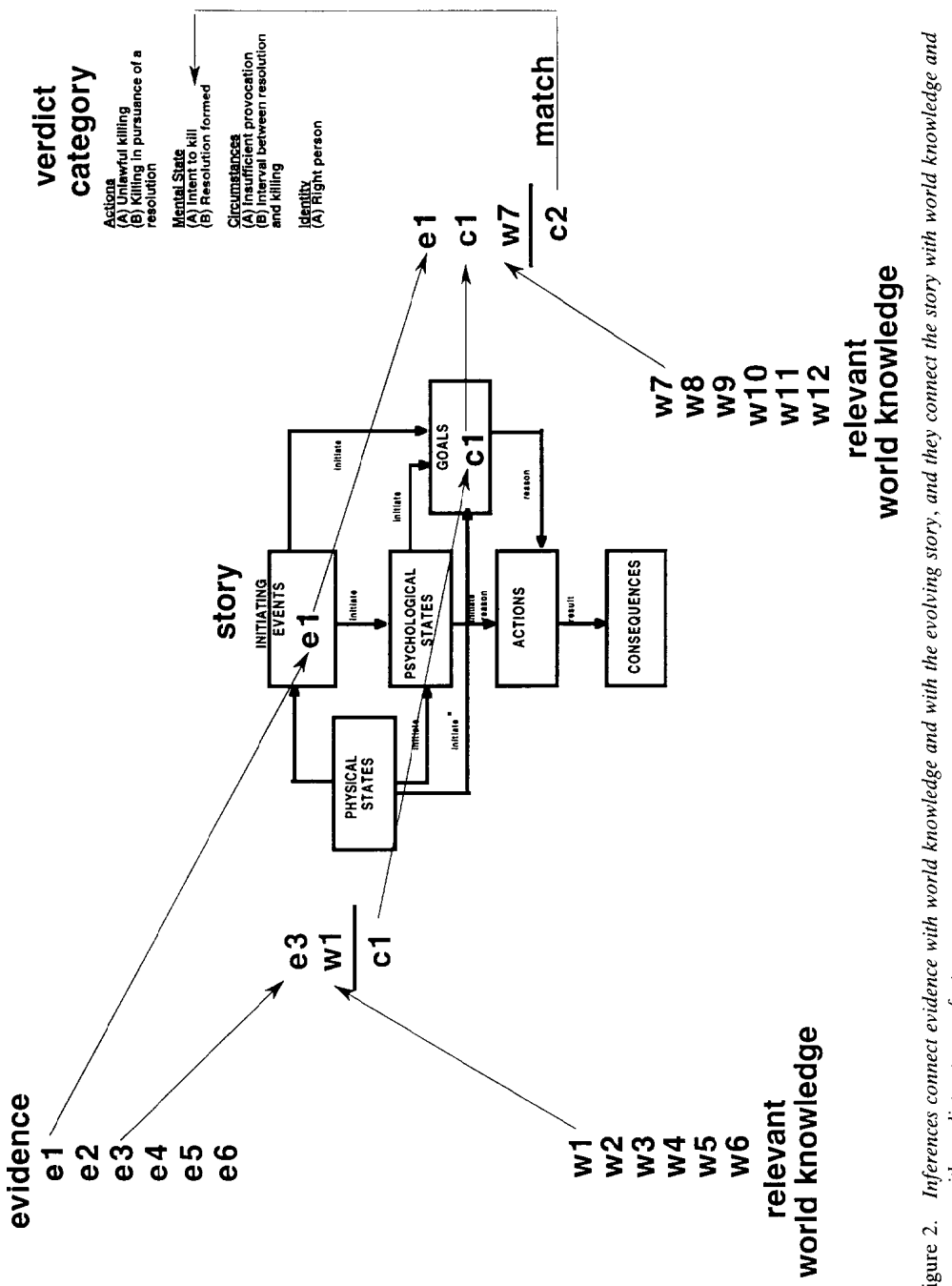

Figure 2. *Inferences connect evidence with world knowledge and with the evolving story, and they connect the story with world knowledge and with verdict category features.*

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). In Fig. 2, we have represented abstractly that a piece of evidence (e3) combined with world knowledge (w1) can produce a conclusion (c1) that becomes an element of the story. This story conclusion might then be involved in a later stage of inferential reasoning required to match the story to a decision alternative. In Fig. 2, the story element (c1) coming from an earlier inference and the story element (e1) coming directly from evidence might be combined with world knowledge (w7) to produce an inferential conclusion allowing those story attributes to be mapped more directly onto the verdict category definition.

In the present research, we have used a taxonomic system to describe inference procedures based on a system for informal reasoning proposed by Allan Collins and his colleagues (Collins, 1978a, 1978b; Collins, Brown, & Larkin, 1980; Collins, Warnock, Aiello, & Miller, 1975). Collins' goal was to develop a system of forms of plausible inference that can be implemented as a well-defined computer program to answer questions based on reasoning from world knowledge in the reasoner's long-term memory. His most recent papers have expressed the system in a formalism consistent with Michalski's variable-valued logical calculus (Collins & Michalski, 1989) and some versions of the system have been implemented as operational programs (Baker, Burnstein, & Collins, 1987; Dontas & Zemankova, 1987; Kelly, 1988).

We have simplified the Collins system and use his general (and traditional) categories of the major inference forms to label the variety of inference types that occur in our analyses of inference in juror decision protocols. In general terms the inference forms include and extend traditional logic inference forms in several ways. The basic inferential forms of deduction, analogy, induction, generalization, and abduction are included in the theory in premise–conclusion form. But the rules of logic are generalized from two values of certainty (true or false) to a continuum of certainty, creating plausible inference forms that would constitute fallacies in logic (e.g., affirming the consequent).

In later sections, in which we discuss our analyses of inference chains in example jurors' reasoning, we outline the inference types and show sample applications in some detail. The theory of reasoning in decision making that we promote has some higher-level claims that we explicate here. First, we lean towards an inclusive definition of the concept of inference, and would include most mental activities that result in adding a new piece of information to the reasoner's current model of a situation. But we would also like to exclude the unconscious, "raw", associative mental activation processes that may underlie many memory retrieval achievements. Although we cannot draw definitive conceptual or operational distinctions between inference processes and other

types of mental processes, we can point to the familiar premise–conclusion propositional form as the prototype of inference processes. Second, although our explanation-based decision-making theory is based on the idea that decision makers construct a model of the situation specific to the decision, our treatment of reasoning processes is based on the idea that we use inference procedures that are fairly general across content domains and contexts in order to construct the model of the situation (cf. Collins, 1978a, 1978b; Collins & Michalski, 1989; Rips, 1986). Thus, the basic inference procedures will be generalizations of syllogistic forms. We also propose, following Collins (1978a, 1978b), that people use several different inference patterns and converging chains of reasoning to establish their conclusions.

Relationship of explanation-based decision making and the story model to other approaches

Our approach has much in common with the "mental models" approaches to deductive reasoning (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991) and discourse comprehension (Bower & Morrow, 1990; Bransford, Barclay, & Franks, 1972; Perrig & Kintsch, 1985; van Dijk & Kintsch, 1983). Like these researchers we propose that people construct mental representations of situations, in the course of normal comprehension and reasoning, that are concrete in the sense that they are constituted of "individuals" with external referents (not abstract "variables") and in that the structure of the representation corresponds to the structure of the relevant situation. Our account of reasoning in decision making is also similar to the process models of deduction and comprehension proposed by these researchers. We claim that a model is constructed, "in the head", of the situation referred to in a problem or text. Additional conclusions are inferred, usually in the form of new relations among elements cited in the original text or problem, and checked for consistency with the source text and world knowledge. We also propose that, sometimes, alternate counter-example models are constructed and evaluated for veracity.

Like the Johnson-Laird and Kintsch representations, the explanation models we have proposed are relatively static, although they usually have a temporal dimension (ordering the events in an alleged crime, a medical case history, or the historical development of a military or economic scenario). But we have not suggested that they can be "run" or "simulated" to project future outcomes. ("Dynamic" mental models have been proposed, but plausible mental procedures have not been spelled out to explain how the models "run". But it seems that sensible "forward inference" procedures could be added to the basic claims about representation to produce the desired "dynamic" models; see Gentner & Stevens, 1983; Kahneman & Tversky, 1982; Rouse & Morris, 1986, for examples.)

The fundamental difference between our explanation-based approach and traditional algebraic approaches, such as cognitive algebra (Anderson, 1981), the lens model (Hammond, Stewart, Brehmer, & Steinman, 1975), and utility theory (von Neumann & Morgenstern, 1947), is that we view *reasoning about the evidence* and *construction of an intermediate, semantic representation of the evidence* to be the central process in decision making, in contrast to an emphasis on the computation that occurs once evidence has been selected, augmented, and evaluated. We also depart from the common assumption that, when causal reasoning is involved in judgment, it can be described by algebraic computations that lead directly to a decision (e.g., Anderson, 1974; Einhorn & Hogarth, 1986; Kelley, 1973). In our model causal reasoning plays a subordinate but critical role by guiding inferences in evidence evaluation and construction of the intermediate explanation.

We can also relate our approach to the cognitive heuristics analysis of judgment under uncertainty (Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1974). We claim that evaluations of uncertainty, estimates of probability, etc., are secondary judgments that occur in the context of primary inference processes. For example, when set membership relationships are involved in inductive, deductive, abductive or analogical inference processes, "representativeness" or the typicality of instances or subsets provides one type of cue about the certainty with which conclusions have been established (see Cherniak, 1984; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; and Rips, 1975, for similar views). Following Collins (1978a, 1978b), we label these secondary considerations "certainty conditions" and postulate additional conditions, beyond the similarity, typicality, ease of recall or generation factors from the Tversky and Kahneman heuristics.

In sum, the basic claim of explanation-based decision making is that constructing an explanation that accounts for evidence (a story in the case of juror decisions) enables critical interpretive processing and organization of the evidence so that evidence can be meaningfully evaluated against multiple judgment dimensions. The story model provides a psychological account for the assignment of relevance to presented and inferred information for juror decisions. Precise claims are made concerning the representational form of the evidence and a mediating role is claimed for stories in subsequent decisions and confidence in those decisions. Uncertainty in the decision is centered in 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 supporting the claims of the story model are provided in other reports (Baltzer & Pennington, 1993; Pennington, 1981; Pennington & Hastie, 1981, 1986, 1988, 1992; Pennington, Messamer, & Nicolich, 1991; Wolfe & Pennington, 1993).

In the present paper we focus on explicating some of the *forms of reasoning*

that occur during the story construction phases of the decision process. In the juror's decision, there is no question that this component dominates the time spent by jurors deciding criminal cases. Our initial approach is descriptive and taxonomic because we believe that it is necessary to have a sample of everyday inferences in hand before it is appropriate to attempt to develop a detailed processing model (see Collins et al., 1975; and Voss, Perkins, & Segal, 1991, for similar approaches).

Research

In previous research, we have established that jurors construct narrative representations of evidence spontaneously in the course of reaching a decision, and that these representations determine their decisions. In the present analyses, our claim is that numerous inferences are made in order to construct and justify these narrative evidence summaries during and after the decision process, that jurors use converging lines of reasoning to establish their major story conclusions, and that the forms of reasoning are well captured by a general theory of plausible reasoning developed by Collins and his colleagues (Collins, 1978a, 1978b; Collins et al., 1975, 1980).

We use Collins' system of categories of major inference forms to label the variety of inference types that occur in juror decision protocols. The basic inference forms we are concerned with involve the mapping of patterns of knowledge onto a conclusion. Four major mapping relations occur frequently in the protocols, as shown in Table 1: deduction, analogy, induction and abduction. A fifth mapping relation, generalization, occurs seldom in this context. These forms vary in terms of whether the mapping is from instance or subset to superordinate category (generalization, abduction), from one set or instance to another (analogy, induction), or from a superordinate set to a subset (deduction).

For each inference mapping shown in Table 1A, there are two forms. The first is a categorical or "set" form in which a property of one set is mapped onto another set. The second is a functional (causal/correlational) form in which the property to be mapped depends on other properties. This dependency is a directional relationship between variables that includes notions of implication, correlation, and causal relationships (see Collins, 1978a).

Collins (1978b) also identifies two causal reasoning principles. One is functional *attribution* (see Table 1B) – a form of induction concerning causes. In its full form, a general causal rule is affirmed (e.g., if the causes $c_1 \ldots c_i$ occur then effect e will follow). When the antecedent events occur and the effect occurs, then conclude that the effect occurred *because of* the causes. A second, functional

alternative is a type of "discounting" reasoning whereby knowledge that one causal antecedent was present, and that an effect occurred, leads to the conclusion that another causal antecedent was probably not present (this form would assume as a premise an appropriate general causal rule). A final principle applies more to sets and properties, *contradiction*, in which knowledge of the presence of one property leads to the inference that contradictory or incompatible values could not also be present.

For each inference form in Collins' theory of plausible reasoning, *certainty conditions* are specified. For example, consider the deduction in Table 1A: P1 = People who frequent bars are loud and impulsive. P2 = Caldwell frequented bars. C = Maybe Caldwell was loud and impulsive. One condition affecting certainty would be how typical (or representative) Caldwell was of people who frequent bars. Thus, this looks like a "representativeness heuristic" judgment under certainty (Tversky & Kahneman, 1983) in that features of a concrete situation (premises 2) is matched against features of a prototypic situation (premise 1), and the certainty of the conclusion is related to the goodness of the match. However, following Collins (1978a, 1978b; Collins et al., 1975) we propose that each inference procedure is associated with appropriate "certainty conditions" (some, but not all, of which have been identified in the literature on judgment under uncertainty; Kahneman, Slovic, & Tversky, 1982). For example, certainty conditions for deductive inferences include: (a) typicality of the instances within the relevant categories (e.g., how typical is Caldwell of those who frequent bars?); (b) base rates of the properties under consideration (e.g., how frequently are people loud and impulsive in general?); (c) variability in properties attributed to general categories (e.g., for the category of people who frequent bars, how variable is the "loud and impulsive" quality?); and (d) dissimilarity between cited instances and the most similar group that behaves in a contradictory manner (e.g., how similar is Caldwell to the kind of person who frequents bars but is *not* loud and impulsive?).

In Collins' (1978a, 1978b) taxonomy of inferences, there are approximately 40 inference forms. In addition to the categorical and functional forms of many of the inferences, many of the mappings also have a *temporal* and *spatial* form in which deductive or abductive mappings are applied to temporal and spatial knowledge. Our goal in this research was to go beyond our previous analyses of the representations of evidence during juror decision making and to characterize the inference structure and lines of argumentation used to coordinate trial information, world knowledge, verdict category information and major story conclusions, in order to produce the representational forms that we have previously analyzed. The 11 inference forms shown in Table 1A and 1B, with a few temporal and spatial variations that are not shown, were adequate to capture most of the juror reasoning we analyzed.

Table 1. *Example inference forms. After Collins (1978a, 1978b)*

Table 1A	Deduction: maps properties of a set onto subsets[a]	Analogy: maps properties from one set onto a similar set	Induction:[b] maps properties of subsets onto other subsets of the same set	Abduction: maps a subset with the same property as a set into the set
Categorical form	Set X has property A1. X1 is subset of X. --- Maybe X1 has A1.	Set X has property A1. Set Y is like set X. --- Maybe Y has A1.	X1, X2 have A1. X1, X2, X3 subsets of X. --- Maybe X3 has A1.	Set X has property A1. X1 has property A1. --- Maybe X1 is subset of X
Example	People who frequent bars are loud and impulsive. Caldwell frequented bars, --- Maybe Caldwell was loud and impulsive.	I am quick tempered. Johnson is probably like me. --- Johnson is probably quick tempered.	Two different friends of mine got in fights in a bar and didn't intend any harm. Johnson got in a fight in a bar. --- Possibly Johnson didn't intend any harm.	Gentle people speak softly. Johnson spoke softly. --- Maybe Johnson was a gentle person.
Functional dependency (causal/correlational) form	A1 depends on B, C. X1 has B, C. --- Maybe X1 has A1.	(A1 depends on B, C.) Y is like X on B, C. X has A1. --- Maybe Y has A1.	(A1 depends on B, C.) X1, X2 have B, C. X1, X2 have A1. X3 has B, C. X1, X2, X3 subsets of X. --- Maybe X3 has A1.	A1 depends on B, C. X1 has A1. --- Maybe X1 has B, C.
Example (positive)	If someone embarrassed Johnson, he would probably go back with malice and do a job on him. Caldwell embarrassed Johnson. --- Probably, Johnson went back with malice to do a job on Caldwell. (modus ponens)	My father carried a knife for protection because he had to come and go in a rough neighborhood. It sounds like Johnson also frequented a rough neighborhood. --- Maybe Johnson carried a knife for protection.	None observed	Going to the park with your family, being home, eating supper, depend on a calm frame of mind. Johnson went to the park with his family, came home, had supper. --- Probably Johnson was in a calm frame of mind.
Example (negative)	If Johnson went back out of pride, he would not need a knife. Johnson took a knife. --- Johnson probably did not go back out of pride. (modus tollens)	None observed	None observed	None observed

Table 1. *continued*

Table 1B	Attribution: for a given subset, when all properties are present, a causal relationship is inferred for the case	Alternative: when two causes are possible and one is known to be true, the other is considered unlikely	Contradiction: when two causes or properties are possible, but contradictory, and one is known to be true, then the other is inferred to not be true
Categorical or functional dependency form	A1 depends on B, C. X1 has B, C. X1 has A1. --- X1 has A1 *because* B, C.	A1 depends on B or C. X1 has A1 and B. --- X1 may not have C.	X has A1. A1 implies not B1. X1 is a subset of X. --- X1 does not have B1.
Example	One reason for carrying a weapon is being afraid. Johnson carried a knife. Johnson was afraid of Caldwell. --- Johnson carried his knife *because* he was afraid of Caldwell.	Johnson went back either out of pride or to fix Caldwell. Johnson went back to fix him. --- Johnson did not go back out of pride.	Johnson was trying to stay away from Caldwel. Staying away from someone is incompatible with trying to find the person. --- Johnson was not trying to find Caldwell.

[a] Subset may mean either several instances or one individual instance.

[b] A fifth standard form, generalization, maps properties of subsets of a set onto the set. This would have the categorical form: X1, X2 have A1; X1, X2 are subsets of X; (conclusion) maybe all X have A1. The functional (causal correlational) form would be: X1, X2 have B, C; X1, X2 have A1; X1, X2 subsets of X; (conclusion) maybe A1 depends on B, C for all X. Generalizations did not occur in the protocols.

Research method

Procedure

Adult subjects were randomly sampled from a pool of more than 200 volunteers who had been called for jury duty in the major trial court in Boston, Massachusetts, and shown a 3-hour videotaped re-enactment of a criminal case (see Pennington & Hastie, 1986, for more details on the method). Following the videotaped trial, the jurors were individually interviewed and asked to talk about their decisions. For present purposes we have selected two jurors' post-decision interviews and will provide detailed analyses of the forms of reasoning and argumentation that occur in their verbal reports.

Subjects

The two example jurors were selected on the basis of their decisions: one chose the most lenient (not guilty) verdict and the other chose the harshest (first-degree murder) verdict. In terms of the lengths and contents of their protocols, they are representative of the other jurors who were interviewed about their decisions. But we must note that the present report relies on two protocols that are being presented to *illustrate* the forms of inference and the roles of inference in the juror decision task. These post-decision interviews contain a mixture of inferences made in the service of the decision and inferences made to communicate or justify the decision. The present method does not allow us to discriminate between these two sources of inferences. Finally, the verbal report interview may omit, even systematically omit, certain classes of difficult-to-report or embarrassing-to-report inferences. We can assert that our current research, including the analysis of verbal reports collected during the presentation of the stimulus trial, replicates the essential richness and variety of inference forms that were obtained in the post-trial presentation method.

Stimulus trial

The specific characteristics of the task will exert a dominant influence on the types and forms of inferences that occur in the decision process. The legal case used in our research, labeled *Commonwealth* v. *Johnson*, represents a typical serious felony trial (it is based on the transcript of a Massachusetts trial) in which the defendant Frank Johnson was charged with killing Alan Caldwell with deliberate premeditation and malice aforethought. We have already briefly

summarized to the stimulus trial, but include a fuller description here in order to highlight the major issues about which subjects would be led to reason.

The first witness for the prosecution, a police officer, testified that one evening from a distance of about 75 feet he saw Caldwell (the victim) hit the defendant Johnson in the face and then Johnson plunged a large knife downward into Caldwell's chest. Because of the distance and the angle of view he was unable to see Caldwell's right hand (to observe whether or not the victim was armed). When arrested, Johnson said, "Caldwell pulled a razor on me so I stuck him".

The second witness, a state pathologist, testified about the cause of death (a stab wound to the heart); that the victim's blood alcohol level rendered him legally drunk; and that the victim's body was marked in several places with scars that may or may not have been surgical in origin. The pathologist also noted that the victim carried a straight razor in his left rear pants pocket.

The owner of the bar outside of which the killing occurred testified that he observed, through a partially obstructed window, and also saw Johnson stab Caldwell. He said that Johnson and a friend (Dennis Clemens) arrived in the bar before Caldwell and that sometime later Caldwell and Johnson left the bar together. The bar owner also reported that earlier on the afternoon of the same day the two men had quarreled and that Caldwell had threatened Johnson with a straight razor and chased him out of the bar. Finally, he noted that Johnson had never caused trouble in the bar before.

The defense opened with testimony from Dennis Clemens, a friend of Johnson's who indicated that he initiated the evening visit to the bar by asking Johnson to join him for a drink. Johnson was reluctant to go into the bar, but when the two men saw no trace of Caldwell, they entered. After they had been in the bar for about half an hour, Caldwell came in and then asked Johnson to step outside with him. Clemens saw the fight start and described Caldwell knocking Johnson to the ground and then drawing a straight razor. Johnson tried to hold Caldwell off (Clemens never saw a weapon in Johnson's hand) and then Caldwell staggered back and fell to the ground.

The second defense witness, a waitress from the bar, corroborated the timing of the fight events and also reported that a car was parked illegally in front of the bar in a location that would have interfered with the police officer's view of the altercation from across the street.

The defendant, Frank Johnson, testified last and said that he had a quarrel with Caldwell in the bar in the afternoon, when a woman asked him (Johnson) for a ride to the dog track. He was frightened and left the bar after Caldwell threatened him with a straight razor. Johnson did not remember having a knife with him at that time; but after going fishing later that day, his wife asked him not to leave his fishing knife around the house where children might find it and he must have put the knife in his pocket. Johnson also reported that he had spent the afternoon with his children at the park.

Later that evening Dennis Clemens invited him to go to the bar again, but he hesitated because he was afraid of Caldwell. After he was convinced that Caldwell was not in the bar, Johnson entered, but later Caldwell came in and sat at the bar. He invited Johnson to step outside, in a friendly way, but then became angry and struck him. The blow knocked Johnson off balance and he was cornered against the wall of the bar; then Caldwell drew his straight razor. Unthinkingly, Johnson drew his fishing knife to protect himself, but Caldwell rushed in and lunged onto the blade, killing himself.

The trial judge's instructions on the law, especially definitions of the four verdicts (choice set alternatives: murder in the first degree, murder in the second degree, manslaughter, and not guilty by reason of self-defense), also exert considerable influence on the jurors' reasoning. For example, to conclude with a verdict of murder in the first degree the juror must be convinced (beyond a reasonable doubt) that a killing was committed with deliberately premeditated malice aforethought. Malice includes feelings of hatred or ill will, but it also includes any intentional infliction of an injury where there is a likelihood of causing death. Malice may be inferred from the intentional use of a deadly weapon without just provocation or legal excuse. In addition, deliberate premeditation is a sequence of thought processes: the plan to murder is formed, then the resolution to kill, then the killing in pursuance of the resolution. On the other end of the spectrum, the juror might find that the killing was in self-defense and the defendant is not guilty of a crime. The right to self-defense arises from a threat to the defendant's life or fear of great bodily harm. The right comes into existence only after the defendant has exhausted all reasonable means to escape from the confrontation and the method of defense can rely on only reasonable force.

In the context of this stimulus trial and the form of the decision prescribed by the judge's instructions, we can anticipate the contents of many of the inferences that will be elicited. The largest subset of the inferences expressed in the protocols were directed at completing a collection of coherent episodes linked by the motivations of the defendant (i.e., to construct a narrative summary of the evidence): What was the defendant's state of mind when he left the bar after the first encounter with the victim in the afternoon? Why did the defendant return to the bar on the evening of the killing? Why did the defendant carry a knife when he returned? How did the fight outside the bar start? Did the defendant attempt to avoid the fatal fight with the decedent? The second most frequent type of inference was directed at questions raised by the legal terms in the judge's instructions: Did the defendant's (inferred) state of mind fit the legal concept of "malice"? Had the defendant "formed a resolution to injure or kill" the victim? Did the defendant "exhaust all means to avoid combat" in the confrontation outside the bar? The third most frequent collection of inferences was concerned with evaluations of the witnesses' credibility: Was the police officer telling the truth? Could the defendant's friend have seen what he claimed to have observed?

Protocol collection and story analysis procedures

Following their exposure to the videotaped trial each juror was instructed: "I would like for you to be a juror in this case and to decide on a verdict. You do not need to decide right away. I would like for you to talk out loud as you think about what to decide." This general solicitation was followed by more specific questions asking jurors if they had considered alternate verdicts, what aspects of the evidence they felt were most important, and what they could recall from the judge's instructions on the law.

The interviews were transcribed (approximately 9,000 words for each juror's responses) and systematic analyses were conducted to extract the summary of the historical events referred to in testimony (the juror's "story") that each juror had constructed from the trial evidence. Detailed reports of these analyses are available elsewhere (Pennington, 1981; Pennington & Hastie, 1986) and we provide diagrammatic summaries of the two example jurors' stories in Figs. 3 and 4.

In Figs. 3 and 4, the stories constructed by these two jurors are shown as the coordination of two streams. The first stream is the event chain showing events that occurred (such as Caldwell threatens Johnson at the bar; Johnson goes home; Johnson goes back to the bar; Johnson takes knife; Caldwell hits Johnson; etc.). This stream looks similar for the two different stories. The second stream represents the interpretation of the event chain or *why* the events occurred as they did; it is the set of goals that explain the events. For example, the first-degree murder juror (Fig. 4) has a goal to stab Caldwell that is the reason for Johnson taking the knife. In contrast, the self-defense juror (Fig. 3) has Johnson taking the knife because he was afraid. It is these goals, internal states, and causal connections between goals and events that are the major story conclusions we propose are derived through the inference procedures that are the topic of our analyses in this report.

Inference analysis procedures

Recall that our claim is that numerous inferences are made to construct and justify these narrative evidence summaries during and after the decision process; that these inferences can be identified according to the general inference forms proposed by Collins (1978a, 1978b); and that inference structures can be identified as lines of reasoning that converge on certain major inferential conclusions. Our first analysis concentrated on identifying the inference *forms* that were used to reach or justify certain story conclusions. These conclusions are concerned with *why the defendant returned to the bar* and *why the defendant was carrying a knife*. Our second analysis concentrated on identifying the *chains* of

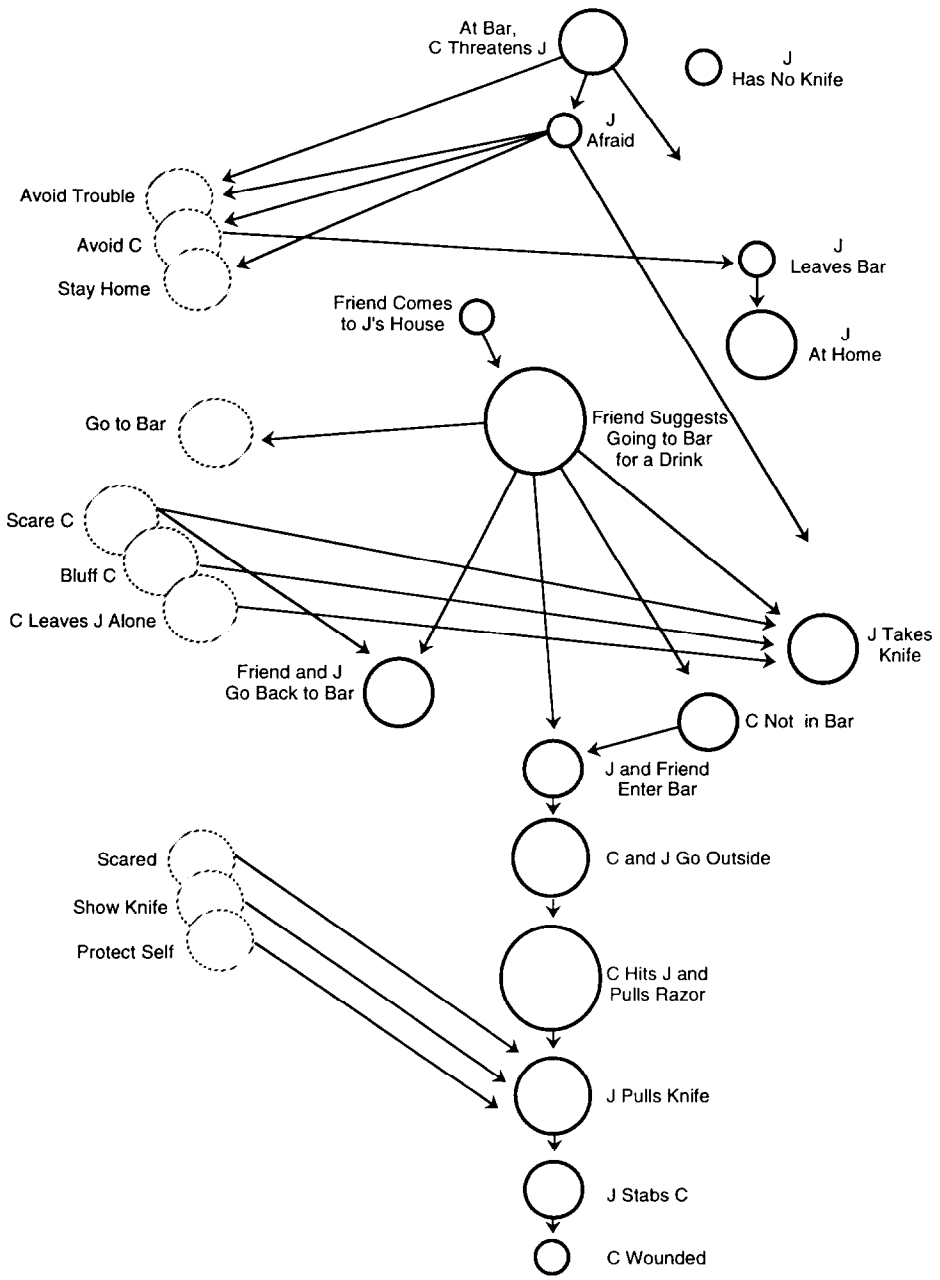

Figure 3. *A summary of causal event chain including goals constructed by juror no. 109 (not-guilty verdict) while making a decision in the Commonwealth* v. *Johnson case. Events and episodes are represented by solid circles and the diameters of the circles indicate the degree of elaboration provided of events by the juror; broken border circles represent the defendant's goals, inferred by the juror. The arrows connect events that were explicitly linked by causal relations in the juror's verbal report. The letters J and C refer to the defendant Johnson and the victim Caldwell respectively.*

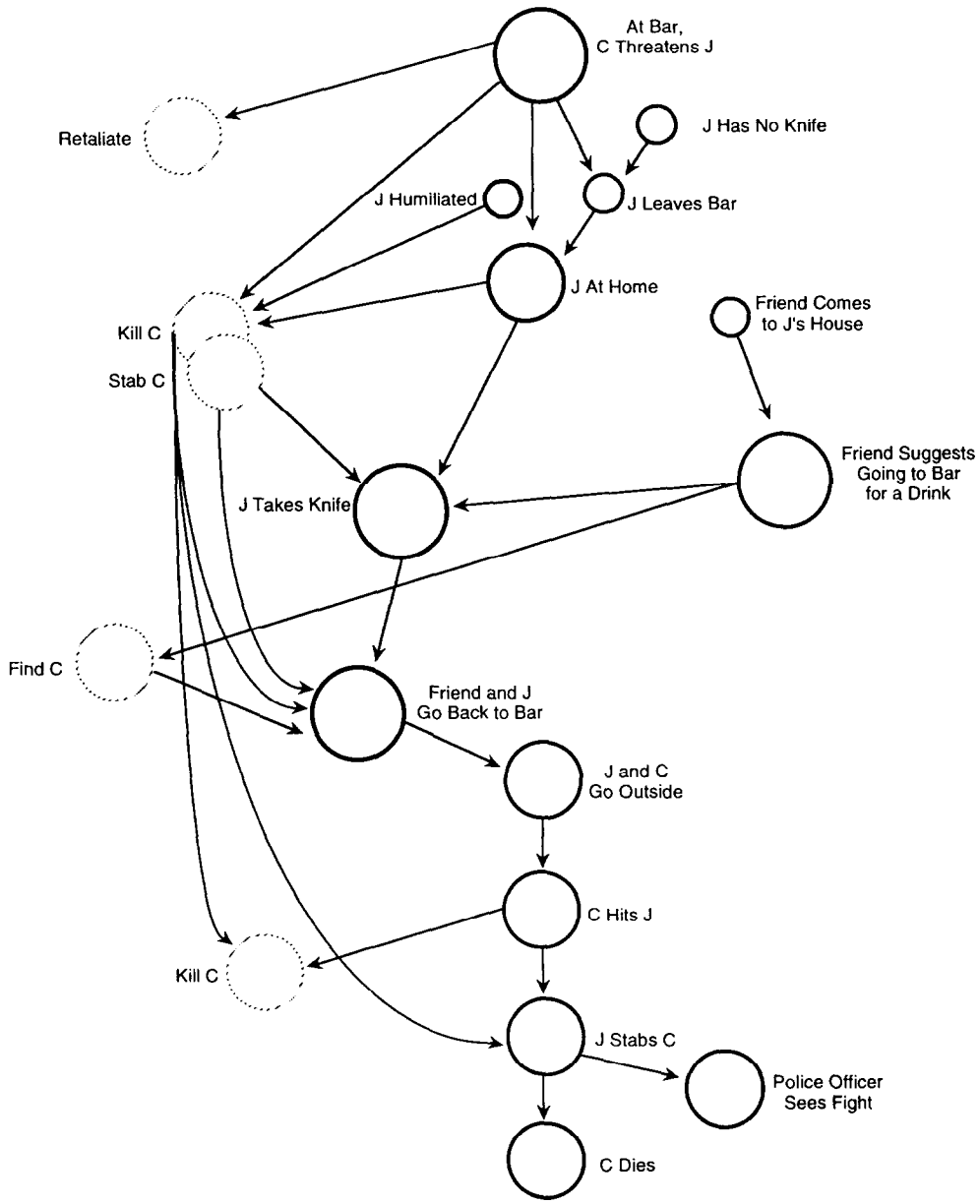

Figure 4. *A summary of causal event chain including goals constructed by juror no. 128 (first-degree murder verdict) while making a decision in the Commonwealth* v. *Johnson case. Events and episodes are represented by solid circles and the diameters of the circles indicate the degree of elaboration provided of events by the juror; broken border circles represent the defendant's goals, inferred by the juror. The arrows connect events that were explicitly linked by causal relations in the juror's verbal report. The letters J and C refer to the defendant Johnson and the victim Caldwell, respectively.*

inferences that were generated by jurors reasoning about the two central conclusions of the stories represented in Figs. 3 and 4. Our final analysis concentrated on identifying the reasoning *strategies* represented by these inference structures.

Tables 2 and 3 list the content and form of the inferences that we identified in each of the two protocols leading to story conclusions concerning why the defendant returned to the bar and why the defendant was carrying a knife. We did this by first identifying each assertion concerning the return to the bar and then identifying all arguments leading to the assertion. We then rewrote the juror's words in the form of an inference consisting of at least two premises and a conclusion.

Jurors don't speak in terms of premises and conclusions. The sentences in the tables are abstracted from the exact words used by jurors during the interview, but we also filled in "missing premises" to make complete inference forms. For example, the not-guilty juror's words concerning the reason Johnson took the knife to the bar were:

> "... I think he took the knife with him because he was scared..."
> "... I think he did it because he was afraid, the gentleman was a big gentleman, did he say 200 pounds, he was well known as a troublemaker..."

These words were translated into two inferences listed as inferences A and B in Table 2. For inference A, the premises "Caldwell was big" and "Caldwell was known to be a troublemaker" were taken from the juror's assertions. However, the premise that "A person who is big and known to be a troublemaker causes people to be afraid" was inferred as the world knowledge premise that allowed the juror to infer that Johnson was afraid of Caldwell. Sometimes jurors would explicitly state such beliefs, but often such beliefs would be implicit in the argument, as in the excerpt above.

A second coding example is drawn from the protocol of the first-degree-murder juror, who said:

> "... That Mr. Johnson, when he came back, he was looking for Mr. Caldwell. Because if somebody had threatened me four or five hours earlier with a straight edge razor, I would not go back to Mr. Gleason's place again, unless I was going to be out for revenge..."
> "... If Mr. Johnson wasn't looking for revenge for this, he would never go back to the same place, even with his friend Mr. Clemens, knowing that Mr. Caldwell had the razor, and he knew he had it from four or five hours ago, if he wasn't going back for revenge and to kill the man..."

These inferences, as coded in our analysis, are shown in Table 3 as inferences

Table 2. *Inferences for not-guilty juror (inference types from Table 1 are shown in parentheses)*

A.	(Functional deduction) A person who is big and known to be a troublemaker causes people to be afraid. Caldwell was big. Caldwell was known to be a troublemaker.	Johnson was afraid of Caldwell.
B.	(Functional attribution) One reason for carrying a weapon is being afraid. Johnson carried a knife. Johnson was afraid of Caldwell.	Johnson took his knife *because* he was afraid of Caldwell.
C.	(Functional abduction) Going to the park with your family, being home, eating supper . . . depend on a calm frame of mind. Johnson went to the park with his family, came home, had supper	Johnson was in a calm frame of mind.
D.	(Contradiction) Johnson was in a calm frame of mind. Being in a calm frame of mind is incompatible with being out for revenge.	Johnson was not out for revenge.
E.	(Negative functional deduction – modus tollens) If Johnson were going home to get his knife for revenge, he would have gone back to the bar immediately. Johnson didn't go back to the bar immediately.	Johnson did not get his knife for revenge.
F.	(Contradiction) Johnson was trying to avoid Caldwell. Avoiding Caldwell is incompatible with seeking revenge.	Johnson was not seeking revenge.
G.	(Categorical deduction) Asthma is associated with nervous problems. Johnson has asthma.	Johnson has nervous problems.
H.	(Contradiction) Johnson had nervous problems. Johnson was afraid of Caldwell. Being nervous and afraid is incompatible with looking for Caldwell.	Johnson was not looking for Caldwell.

continued

Table 2. *continued*

I.	(Self-analogy) If someone were going to fight with me, I wouldn't want to go anywhere near them. Johnson is like me. --- If someone were going to fight with Johnson, he wouldn't want to go anywhere near them.
J.	(Functional deduction – modus ponens) If someone were going to fight with Johnson, he wouldn't want to go anywhere near them. Caldwell wanted to fight with Johnson. --- Johnson did not want to go anywhere near Caldwell.
K.	(Contradiction) Johnson was trying to stay away from Caldwell. Staying away from someone is incompatible with planning to find them. --- Johnson didn't plan and think about finding Caldwell.
L.	(Abduction) A favorite place is where you go all the time. They drink at Gleason's all the time. --- Gleason's is their favorite place.
M.	(Functional deduction – modus ponens) People like to go to their favorite place for relaxation. Gleason's is Johnson's favorite place. --- Johnson likes to go to Gleason's.
N.	(Functional attribution) Going somewhere depends on liking the place and being invited. Clemens asked Johnson to go to Gleason's. Johnson likes to go to Gleason's. Johnson went to Gleason's. --- Johnson went to Gleason's *because* Clemens asked and Johnson likes to go to Gleason's.

A, B, C, and D. Again, many world knowledge premises must be inferred by the coders to complete the inference forms. Although our analyses were highly inferential, two coders independently applied similar coding methods to identify the inferences in Tables 2 and 3. For these inferences and others in the two protocols, there was substantial agreement (over 80%) on the contents and forms of the inferences abstracted by the two coders.

One question we posed in these analyses was the extent to which we could

Table 3. *Inferences generated by first-degree murder juror (inference types from Table 1 are shown in parentheses)*

A.	(Self-analogy) If somebody threatened me with a razor, I would not go back to the same place unless I were looking for revenge. Johnson is like me.	If somebody threatened Johnson with a razor, he would never go back to the same place unless he were looking for revenge.
B.	(Functional deduction – modus ponens) If somebody threatened Johnson with a razor, he would never go back to the same place unless he were looking for revenge. Caldwell threatened Johnson with a razor. Johnson returned to the same place.	Johnson was looking for revenge.
C.	(Negative functional deduction – modus tollens) If Johnson wasn't looking for revenge, he would never go back to the same place. Johnson went back to the same place.	Johnson was looking for revenge.
D.	(Temporal analogy) Johnson knew Caldwell had a razor on him in the afternoon. The evening is near in time to the afternoon.	Johnson knew that Caldwell would have a razor in the evening.
E.	(Deduction) A friend is not much help when confronted by a razor. Johnson knew Caldwell would have a razor.	Johnson's friend would not be much help.
F.	(Deduction) Places where people have razors are dangerous. Johnson knew that Caldwell would have a razor in the evening.	Gleason's would be dangerous in the evening.
G.	(Functional attribution) One reason for returning to a place known to be dangerous without much help is to get revenge. Gleason's would be dangerous in the evening. Clemens would not be much help. Johnson returned to Gleason's. Johnson was looking for revenge.	Johnson returned to Gleason's *because* he wanted revenge.
H.	(Abduction) A renitent person says things like... Johnson (in testimony) said things like...	Johnson is a renitent person (doesn't like to back down).

continued

Table 3. *continued*

I.	(Negative functional deduction – modus tollens) Having to back down makes you an underdog. Johnson didn't want to be an underdog.
	Johnson wasn't going to back down.
J.	(Functional abduction) Not backing down depends on going back and showing Caldwell who is boss by killing him. Johnson is not going to back down.
	Johnson went back to show Caldwell who is boss by killing him.
K.	(Self-analogy) If anybody pulled a razor on me in a bar, I would never go back unless I was going to retaliate with another weapon. Johnson is like me.
	If anybody pulled a razor on Johnson in a bar, he would never go back unless he was going to retaliate with another weapon.
L.	(Functional attribution) If anybody pulled a razor on Johnson in a bar, he would never go back unless he was going to retaliate with another weapon. Caldwell pulled a razor on Johnson. Johnson went back with a knife.
	Johnson went back with a knife *because* he was going to retaliate with a weapon.
M.	(Negative functional deduction – modus tollens) If Johnson went back out of pride, he would not need a knife. Johnson took a knife.
	Johnson did not go back out of pride.
N.	(Functional alternative) Johnson went back either out of pride or to fix Caldwell. Johnson went back to fix him.
	Johnson did not go back out of pride.
O.	(Self-analogy) If someone embarrassed me, I would go back with malice and do a job on him. Johnson is like me.
	If someone embarrassed Johnson, he would probably go back with malice and do a job on him.
P.	(Functional deduction – modus ponens) If someone embarrassed Johnson, he would probably go back with malice and do a job on him. Caldwell embarrassed Johnson.
	Probably, Johnson went back with malice to do a job on Caldwell.

successfully capture the inferential reasoning of the jurors by the inference forms proposed by Collins. We have shown the forms we used most frequently in Table 1, along with examples. A typical *deduction* from world knowledge (Table 2, inference A) consists of the following three premises and conclusion from the not-guilty juror's protocol:

Premise 1: Reasons to be afraid of someone include if that person is a troublemaker and is big. (source: world knowledge)
Premise 2: Johnson new Caldwell was a troublemaker. (source: trial testimony)
Premise 3: Caldwell was big. (source: trial testimony)

Conclusion: Johnson was afraid of Caldwell.

Typically, for central decision-relevant conclusions several lines of inference were attempted and their success or failure contributed to the juror's confidence in the common conclusion. Thus, the juror who initially concluded "Johnson was afraid" develops another line of support, in the form of an *analogy* to himself, followed by a deduction, for the conclusion about Johnson's motivational state (Table 2, inferences I and J):

Premise 1: If someone were going to fight with me, I would not go anywhere near them. (source: personal knowledge)
Premise 2: Johnson is like me. (source: matching to personal knowledge)

Conclusion: If someone were going to fight with Johnson, *he* wouldn't go anywhere near them.

Premise 1: If someone were going to fight with Johnson, he wouldn't go anywhere near them. (source: prior inference)
Premise 2: Caldwell wanted to fight with Johnson. (source: trial testimony)

Conclusion: Johnson did not want to go anywhere near Caldwell.

There is also a substantial amount of reasoning by *contradiction*, especially by the juror who concluded with the not guilty verdict (Table 2, inference D):

Premise 1: Johnson was in a calm state of mind. (source: prior inference)
Premise 2: Being in a calm state of mind is incompatible with being out for revenge. (source: general world knowledge)

Conclusion: Johnson was *not* out for revenge.

There was also a substantial amount of reasoning by *negative deduction* (or modus tollens, Table 2, inference E):

Premise 1: If Johnson were going home to get his knife for revenge, he would have gone back to the bar immediately (source: general world knowledge)
Premise 2: Johnson did not go back immediately. (source: trial testimony)

Conclusion: Johnson was *not* out for revenge.

Comparable examples of the same inference forms can be found in the protocol of the juror who concluded with the contrasting decision, first-degree murder. For example, a self-analogy form followed by a deduction supports the conclusion that Johnson was motivated by malice when he returned to the bar (Table 3, inferences O and P):

Premise 1: If someone embarrassed me, I would go back with malice and do a job on him. (source: personal knowledge)
Premise 2: Johnson is like me. (source: matching to personal knowledge)

Conclusion: If someone embarrassed Johnson, he would probably go back with malice and do a job on him.

Premise 1: If someone embarrassed Johnson, I would go back with malice and do a job on him. (source: prior inference)
Premise 2: Caldwell embarrassed Johnson. (source: prior inferences)

Conclusion: Probably, Johnson went back with malice to do a job on Caldwell.

For our second analysis, we again focused on the conclusions, for the not-guilty juror: *the defendant returned to the bar on the evening of the killing because his friend invited him and he likes to drink in that locale* (Table 2, inference N) and *he carried a knife because he was afraid* (Table 2, inference B). And for the juror who found the defendant guilty of first-degree murder, we focused on the parallel inferences relevant to his conclusion that *the defendant returned to the bar on the evening of the killing because he wanted revenge* (Table 3, inferences G and P) and the conclusion that *he carried a knife because he intended to assault and kill the victim* (Table 3, inferences L and P). In this analysis, we linked inferences together in chains any time a prior inferential conclusion appeared as one of the premises in a subsequent inference. Inference chains are shown in Fig. 5 for the not-guilty juror and in Fig. 6 for the first-degree murder juror.

Some general properties of everyday inferences are suggested by diagrams of the inference chains leading to central conclusions (Figs. 5 and 6). First, at least under conditions where a reasoner is motivated to think hard about conclusions, we see the tendency to create multiple independent inference chains relevant to central conclusions. Thus, in the present example, two to four chains are generated for each of the major decision-relevant conclusions concerning the

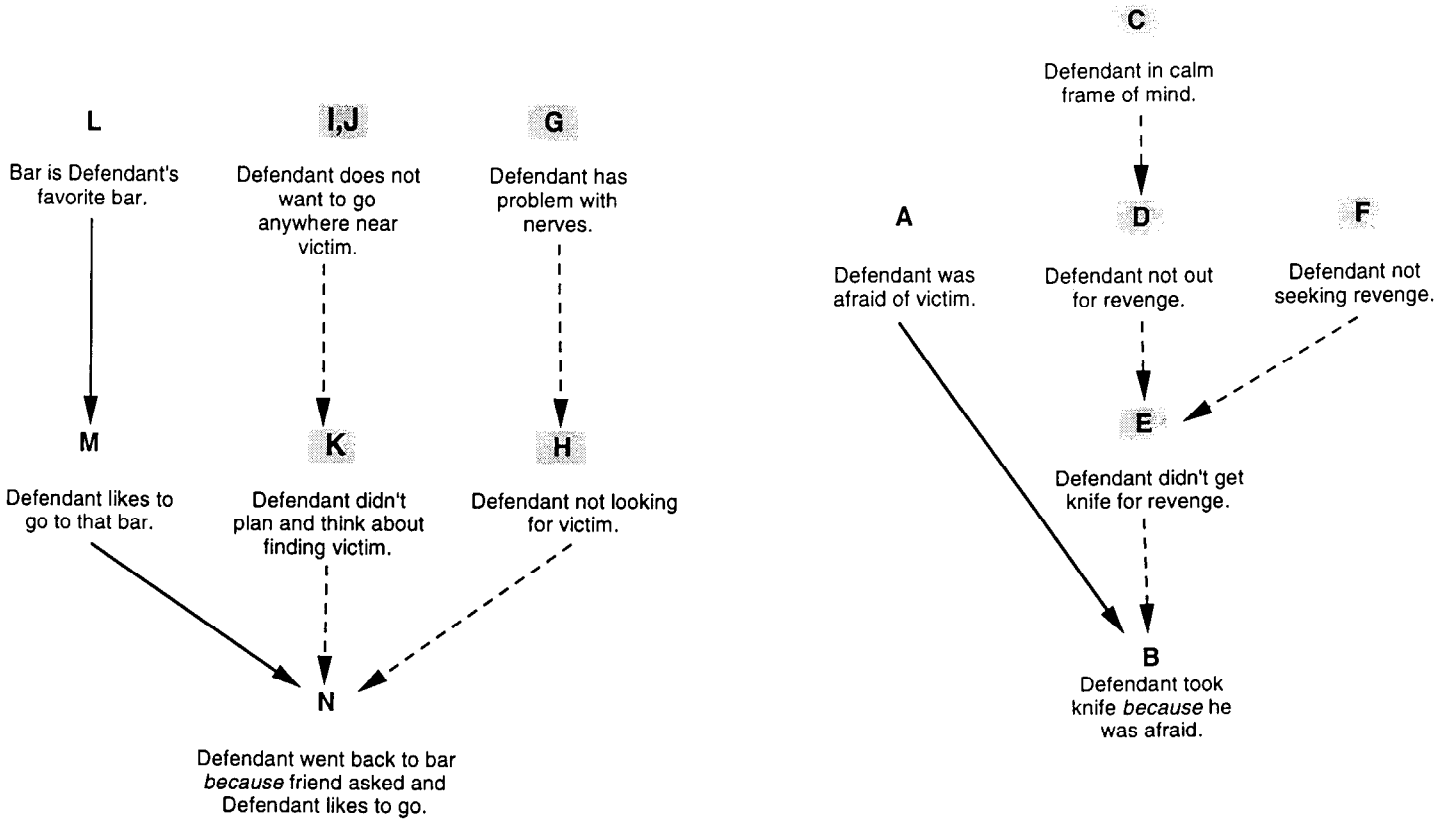

Figure 5. *Example inference structures for juror no. 109 (not-guilty verdict), underlying conclusions concerning the defendant's motivations to return to the bar and to carry a knife in the evening of the killing. Only inference conclusions are listed. Key: Letters A–N refer to inferential conclusions from Table 2. The two conclusions (B and N) correspond to episodes on the causal chain maps (Fig. 4). A solid line connecting inferences indicates a directly confirming argument. The dotted line and the shaded letters indicate an argument for a position by disconfirming opposing arguments. Collins' inference types are not fully illustrated here since the premises for each inference are not shown (see Table 2).*

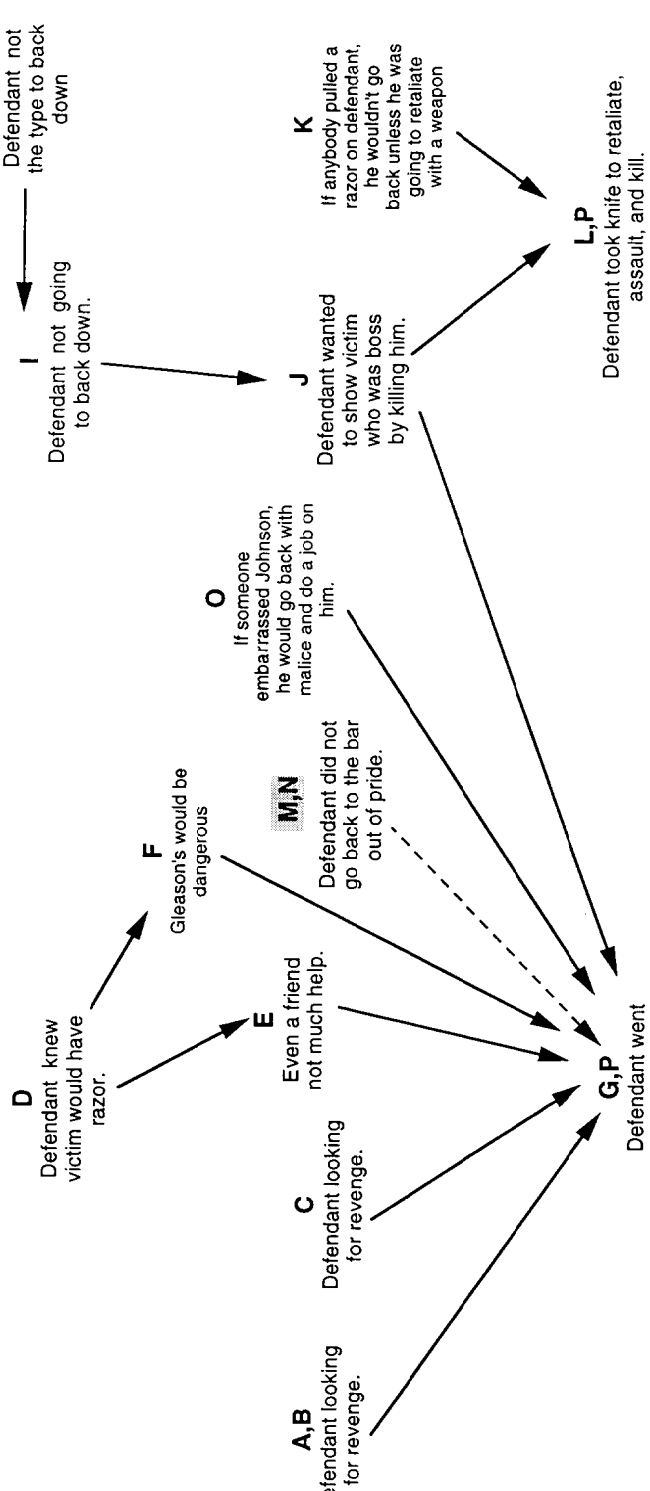

Figure 6. *Example inference structures for juror no. 128 (first-degree murder verdict), underlying conclusions concerning the defendant's motivations to return to the bar and to carry a knife on the evening of the killing. Only inference conclusions are listed. Key: Letters A–P refer to inferential conclusions from Table 3. The two conclusions (G, P and L, P) correspond to episodes on the causal chain maps (Fig. 3). A solid line connecting inferences indicates a directly confirming argument. The dotted line and the shaded letters indicate an argument for a position by disconfirming opposing arguments. Collins' inference types are not fully illustrated here since the premises for each inference are not shown (see Table 3).*

defendant's motives for returning to the bar and for carrying a weapon. The chains are also relatively short; the longest example contains three inferential steps.

Differences in strategies can also be detected from the inference structures created by the linked inferences. In Fig. 5, the not-guilty juror displays a single confirmatory argument (arguing *for* her conclusion) and two disconfirming arguments (arguing *against* competing conclusions, indicated in the figure by dotted lines). In contrast, the first-degree murder juror, shown in Fig. 6, uses converging confirmatory arguments almost exclusively. In addition, this juror relies much more heavily on the self-analogy strategy in which inferences are based on imagining himself in the place of the defendant.

Our analyses show that explanation-based decision makers' thought processes are full of connected chains of propositions that can be reliably classified into the categories of inference forms provided by Collins' system of plausible inference. Our current research shows that the wealth of inferences revealed in the post-trial interviews reported here is also present in on-line verbal reports, elicited after each witness during the trial. Evidence for the utility of a system like Collins' is provided by the finding that only a few general inference forms suffice to classify a large portion of the inferences in the protocols; the same forms keep reappearing within and between protocols, regardless of the ultimate verdict decision reached by the juror. The links between elements in the chains are more than simple semantic associations; rather they repeatedly fit the premise–conclusion categories prescribed by Collins' system.

In the context of the juror's decision task and the particular case materials, certain interesting inference forms frequently appear: (a) reasoning in terms of "negation" (by contradiction, and by negative deduction (modus tollens)) is surprisingly common; (b) reasoning by analogy to (hypothetical) self behavior is frequent and appears to be a very common basis for inferences of social motives (cf. Dawes, 1990; Ross, Greene, & House, 1977); and (c) there is extensive "discounting" of alternate causes ("functional alternative" in Collins' terms; Kelley, 1973).

Local confidence

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 or heuristic calculations could be incorporated into our theory to yield the level of certainty with which a juror believes in any particular proposition (and consequently in the ultimate decision proposition). As we have already noted, deductive inferences such as inference A from

Table 2 looks like a "representativeness heuristic" judgment under uncertainty (Tversky & Kahneman, 1983) in that features of a concrete situation (premises 2 and 3) are matched against features of a prototypic situation (premise 1), and the certainty of the conclusion is related to the goodness of the match. We have proposed, following Collins (1978a, 1978b; Collins et al., 1975) that each inference procedure is associated with additional certainty conditions that will determine local confidence in inferential conclusions. For example, certainty conditions for deductive inferences include: (a) typicality of the instances within the relevant categories (e.g., How typical a troublemaker is Caldwell? How big is Caldwell? How typical a person is Johnson?); (b) base rates of the properties under consideration (e.g., How frequently are people afraid of other people?); (c) variability in properties attributed to general categories (e.g., What is the range of emotional reactions to big troublemakers?); and (d) dissimilarity between cited instances and the most similar group that behaves in a contradictory manner (e.g., How similar is Johnson to the kind of person who would *not* be afraid of a big troublemaker?).

Because of lack of empirical support for Bayesian calculations as a *description* of human judgment under uncertainty, and because of the incompleteness of the heuristic approach, 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. Our main assumption is that at the time that an inferential conclusion is being considered as a potential story event, certainty evaluations render it certainly true (and therefore as data, e.g., Premise 2), or as uncertain (and therefore as a hypothesis, e.g., premise 3, conclusion), or as rejected and therefore certainly untrue. The final level of acceptability of any given proposition is hypothesized to be a function of its inferential support (with reference to certainty conditions and the evaluation of contradictory conclusions), its plausibility with reference to known "facts" about the world, and its role in the larger story structure.

We do not deny the significance of confidence or probability evaluation processes; but we believe that their role is secondary and embedded in logic-like inference procedures. One sign of the secondary role of these factors is their shadowy appearance in think-aloud reports of judgment under uncertainty. For example, in the empirical analyses we report, discussion of degrees of certainty, confidence, or probability accounts for a small percentage (less than 5%) of the verbal report protocols.

Global confidence

We have already outlined features of the semantic representation of decision-relevant evidence that we believe contribute to global confidence in the decision.

According to our theory, two certainty principles govern acceptance of a story: *coverage* and *coherence*. An additional certainty principle, *uniqueness*, will contribute to confidence.

A very important direction for development of the story model and for explanation-based decision making involves elaborating and formalizing the principles that we suggest determine confidence in decisions: coverage, coherence (completeness, consistency, and plausibility), uniqueness, and goodness-of-fit. One part of the puzzle involves examining these principles separately and in interaction. For example, in actual case materials, the coverage of the story and its completeness would often be correlated, although in principle they need not be. Experiments can examine the effects of these variables independently. Our empirical work so far has suggested that there will be interactions we have not anticipated. For example, we suggested that uniqueness (the extent to which there is only one coherent story) will enhance confidence. Yet, we found that in one set of case materials, mock-jurors were most confident when they were able to easily construct both stories (Pennington & Hastie, 1988). We suspect this was because one of the stories was less plausible (and thus less coherent) than the other. On the other hand, it could be that knowing both stories will always increase confidence, regardless of plausibility. These are clearly empirical questions that are currently being addressed through experimentation (Pennington et al., 1991).

A second goal in pursuing determinants of confidence is 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. Toward this end, we have adopted a formalization of the goodness-of-fit, coverage, coherence, and uniqueness ideas, based on an adaptation of Thagard's model of explanatory coherence, ECHO (Thagard, 1989). We call our modification of Thagard's (1989) explanatory coherence model STORY-ECHO, for the obvious reason that we hypothesize that explanations of the evidence in legal trials take the form of stories (see Pennington et al., 1991, for preliminary reports of this work). The model is an interactive activation model that represents hypotheses and supporting evidence in a network of interconnected propositions (Rumelhart & McClelland, 1986). Similar models have been implemented to describe comprehension processes and representations in non-decision-making tasks (e.g., Kintsch, 1988) and proposed for analogical and deductive reasoning tasks that are similar to our legal judgment task (Holyoak & Thagard, 1989). Although our modeling work is preliminary, we expect to be able to provide more explicit comparisons between our ideas about confidence and uncertainty and those proposed by probability calculi as applied to the legal decision task (see Thagard, 1991, for such a comparison in the area of belief in scientific theories).

Concluding comments

Our primary goal in this paper was to delineate the roles of inference procedures in explanation-based decision processes with illustrations of the pervasiveness and variety of inferences in this type of decision. Considerable psychological research has focused on the question of the generality of inference rules or procedures (Smith, Langston, & Nisbett, 1992). Are very general inference rules, like those represented in the system we applied in our analysis, sufficient to account for most everyday reasoning (Johnson-Laird, 1983; Rips, 1983, 1990)? Or must we postulate intermediate-level domain-specific schemas (Cheng & Holyoak, 1985; Cosmides, 1989; Gigerenzer & Hug, 1992) or even situation-specific rules or case-based procedures (Evans, 1989; Kolodner, 1992; Pollard, 1982) to account for everyday reasoning?

Although we cannot claim to have shown that less general pragmatic schemas or concrete situation-based rules are not relevant, we have been able to account for a substantial portion of the contents of our protocols with the general procedures in Collins' system. The contents of premises and conclusions do matter (e.g., we predict they will have large effects on the confidence with which conclusions are asserted), but the basic forms of reasoning seem always to fit into Collins' forms. The one possible elaboration of Collins' system that we might propose would be to attribute a special status to the "analogy to self" form that appears to be ubiquitous in social inference processes.

We also want to emphasize that in the context of explanation-based decision making, for the types of general rules and knowledge structures we have proposed, some of the disputes between proponents of general inference rules and proponents of mental models (Johnson-Laird, 1983, and Johnson-Laird & Byrne, 1990; vs. Rips, 1986, 1989b) are resolved by embracing both points of view in one framework. Our claim is that general inference rules are the primary tools for the construction, interpretation, and application of an explanatory mental model to perform an overarching decision task. Thus, we suggest that there are complementary roles for inference rules (*construction* of a model of the judgment-relevant situation) and situation models (*representation* of the situation implied by the judgment-relevant evidence).

If general inference rules and situation-specific explanatory models are involved in a cooperative process, we need to further specify the roles of these components. One advantage of the combination of rules and models in the explanation-based framework is that it suggests a solution to the frequently raised issue of the unconstrained generativity of simple forms of the inference rule and mental model approaches. Our claim is that the decision maker's goal is to construct a complete representation of the judgment-relevant situation that "covers" the available valid evidence. In a sense, this is a specification of some "extra-logical constraints", to indicate which of the many possible models will be

constructed and, indirectly, which inferences are likeliest to be drawn (cf. Johnson-Laird & Byrne, 1991).

We also need to say something about the types of rules or operators that can be applied to an explanation model once it is constructed. In the juror decision, we have concluded that the primary operations are feature comparisons to determine if the constructed story fits any of the verdict categories. However, we would also like to allow inference rules to take "parts" of the model as premises for forward inferences. For example, many jurors in the Johnson case reason that if Johnson has five children and Johnson is convicted and incarcerated it will mean a substantial hardship for his family. At present, without more empirical research to illuminate these processes, we can only indicate where in the larger framework the relevant specifications would occur.

A secondary goal of the paper is to present additional detailed examples of explanation-based decision making. However, we do not claim that explanation-based decision making is the only decision-making strategy available to decision makers, nor do we claim it will be applied everywhere. Elsewhere we have shown that even in the context of a juror decision task other general models (i.e., anchor and adjust inductive processes) can be induced by appropriate instructions and response requirements (Pennington & Hastie, 1992). Furthermore, in many popular laboratory research tasks where a decision is based on a relatively small set of independent evidence items, where the required judgment dimension is unidimensional and known prior to hearing evidence, or where the to-be-judged outcome is "caused" by a random process (e.g., where subjects are asked to assess the attractiveness of lottery gambles or to make repeated evaluations of personality descriptions), it is difficult to see why a subject would be motivated to construct a complex causal model of the relevant situation. Under such conditions, we believe that algebraic models such as those based on linear additive, anchor-and-adjust updating processes provide an adequate picture of the judgment strategy (Anderson, 1981; Einhorn & Hogarth, 1986; Hammond et al., 1975; Lopes, 1982).

However, for decisions in which a complex base of interdependent evidence items is considered and in which the features of the ultimate decision are not all known prior to evaluation of the evidence, we would expect the explanation-based model to describe behavior and expect to observe the rich display of plausible reasoning processes we have found in legal decisions.

Acknowledgements

Support for this research was provided by NSF Grant SES-9113479 to the first author and by NSF Grant BNS-8717259 to the second author.

References

Anderson, N.H. (1974). Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 1–101). New York: Academic Press.

Anderson, N.H. (1981). *Foundations of information integration theory*. New York: Academic Press.

Axelrod, R. (Ed.) (1976). *Structure of decision: The cognitive maps of political elites*. Princeton: Princeton University Press.

Baker, M., Burnstein, M.H., & Collins, A. (1987). Implementing a model of human plausible reasoning. In *Proceedings of the Tenth International Joint Conference of Artificial Intelligence* (Vol. 1, pp. 185–188). Los Altos, CA: Morgan Kaufman.

Baltzer, A.J., & Pennington, N. (1983). *Reasoning about conjunctions and disjunctions of events: An explanation-based account*. Unpublished manuscript, Psychology Department, University of Colorado, Boulder.

Bennett, W.L., & Feldman, M. (1981). *Reconstructing reality in the courtroom*. New Brunswick, NJ: Rutgers University Press.

Bostrom, A., Fischhoff, B., & Morgan, M.G. (1992). Characterizing mental models of hazardous processes. A methodology and an application to radon. *Journal of Social Issues*, *48*, 85–100.

Bower, G.H., & Morrow, D.G. (1990). Mental models in narrative comprehension. *Science*, *247*, 44–48.

Bransford, J.D., Barclay, J.R., & Franks, J.J. (1992). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, *3*, 193–209.

Cheng, P.W., & Holyoak, K.J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*, 391–416.

Cherniak, C. (1984). Prototypicality and deductive reasoning. *Journal of Verbal Learning and Verbal Behavior*, *23*, 625–642.

Collins, A. (1978a). Fragments of a theory of human plausible reasoning. In D. Waltz (Ed.), *Theoretical issues in natural language processing II* (pp. 194–201). Urbana, IL: University of Illinois Press.

Collins, A. (1978b). *Human plausible reasoning* (Report No. 3810). Cambridge, MA: Bolt, Beranek & Newman Inc.

Collins, A., Brown, J.S., & Larkin, K.M. (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.

Collins, A., Warnock, E.H., Aiello, N., & Miller, M.L. (1975). Reasoning from incomplete knowledge. In D. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 383–415). New York: Academic Press.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? *Cognition*, *31*, 187–276.

Crothers, E.J. (1979). *Paragraph structure inference*. Norwood, NJ: Ablex.

Dawes, R.M. (1990). The potential nonfalsity of the false consensus effect. In R.M. Hogarth (Ed.), *Insights in decision making* (pp. 179–199). Chicago: University of Chicago Press.

de Kleer, J., & Brown, J.S. (1983). Assumptions and ambiguities in mechanistic mental models. In D. Gentner & A.L. Stevens (Eds.), *Mental models* (pp. 155–190). Hillsdale, NJ: Erlbaum.

Dontas, K., & Zemankova, M. (1987). APPLAUS: An implementation of the Collins–Michalski theory of plausible reasoning. In *Proceedings of the Third International Symposium on Methodologies for Intelligent Systems*, Torino, Italy, 1987.

Einhorn, H.J., & Hogarth, R.M. (1986). Judging probable cause. *Psychological Bulletin*, *99*, 3–19.

Elwork, E., Sales, B.D., & Alfini, J.J. (1977). Juridic decisions: In ignorance of the law or in light of it? *Law and Human Behavior*, *1*, 163–189.

Evans, J.St.B.T. (1989). *Bias in human reasoning: Causes and consequences*. Hove, UK: Erlbaum.

Gentner, D., & Stevens, A.L. (Eds.) (1983). *Mental models*. Hillsdale, NJ: Erlbaum.

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition*, *43*, 127–171.

Hammond, K.R., Stewart, T.W., Brehmer, B., & Steinman, D. (1975). Social judgment theory. In M. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes*. New York: Academic Press.

Hastie, R., Penrod, S.D., & Pennington, N. (1983). *Inside the jury*. Cambridge, MA: Harvard University Press.

Hogarth, R.M., Michaud, C., & Mery, J.L. (1980). Decision behavior in urban development: A methodological approach and substantive considerations. *Acta Psychologica*, *45*, 95–117.

Holyoak, K.J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, *13*, 295–356.

Hutchins, E. (1980). *Culture and inference*. Cambridge, MA: Harvard University Press.

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Johnson-Laird, P.N., & Byrne, R.M.J. (1990). Meta-logical puzzles: Knights, knaves, and Rips. *Cognition*, *36*, 69–84.

Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction*. Hove, UK: Erlbaum.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–210). New York: Cambridge University Press.

Kaplan, J. (1978). *Criminal justice: Introductory cases and materials* (2nd Edn.). Mineola, NY: Foundation Press.

Kelley, H.H. (1973). The processes of causal attribution. *American Psychologist*, *28*, 107–128.

Kelly, J. (1988). *PRS: A system for plausible reasoning*. Master's thesis, Computer Science Department, University of Illinois, Champaign.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182.

Klein, G.A., Orasanu, J., Calderwood, R., & Zsambok, C.E. (Eds.) (1993). *Decision making in action: Models and methods*. Norwood, N.J.: Ablex.

Kolodner, J. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review*, *6*, 3–34.

Lewis, C.H. (1988). Why and how to learn why: Analysis-based generalization of procedures. *Cognitive Science*, *12*, 211–256.

Lopes, L.L. (1982). *Toward a procedural theory of judgment*. Technical Report No. 17, Wisconsin Human Information Processing Program, Madison.

Mandler, J.M. (1984). *Stories, scripts, and scenes: Aspects of schema theory*. Hillsdale, NJ: Erlbaum.

Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289–316.

Osherson, D.N., Smith, E.E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185–200.

Pennington, N. (1981). *Causal reasoning and decision making: The case of juror decisions*. Unpublished doctoral dissertation, Harvard University.

Pennington, N., & Hastie, R. (1981). Juror decision making models: The generalization gap. *Psychological Bulletin*, *89*, 246–287.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, *51*, 242–258.

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: The effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 521–533.

Pennington, N., & Hastie, R. (1991). A cognitive theory of juror decision making: The story model. *Cardozo Law Review*, *13*, 519–557.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology*, *62*, 189–206.

Pennington, N., & Hastie, R. (1993). A theory of explanation-based decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 188–204). Norwood, NJ: Ablex.

Pennington, N., Messamer, P.J., & Nicolich, R. (1991). *Explanatory coherence in legal decision making*. Unpublished manuscript, Psychology Department, University of Colorado, Boulder.

Perrig, W., & Kintsch, W. (1985). Propositional and situational models of text. *Journal of Memory and Language*, *24*, 503–518.

Pollard, P. (1982). Human reasoning: Some possible effects of availability. *Cognition*, *12*, 65–96.

Pople, H.E., Jr. (1982). Heuristic methods for imposing structure on ill-structured problems: The structuring of medical diagnostics. In P. Szolovits (Ed.), *Artificial intelligence in medicine* (pp. 119–190). Boulder, CO: Westview Press.

Rips, L.J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, *14*, 665–681.

Rips, L.J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, *90*, 38–71.

Rips, L.J. (1986). Mental muddles. In M. Brand & R.M. Harnish (Eds.), *The representation of knowledge and belief* (pp. 258–286). Tucson: University of Arizona Press.

Rips, L.J. (1989a). Similarity, typicality, and categorization. In S. Vosniadu & A. Ortony (Eds.), *Similarity and analogy* (pp. 21–59). Cambridge, UK: Cambridge University Press.

Rips, L.J. (1989b). The psychology of knights and knaves. *Cognition*, *31*, 85–116.

Rips, L.J. (1990). Reasoning, *Annual Review of Psychology*, *41*, 321–353.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution process. *Journal of Experimental Social Psychology*, *13*, 279–301.

Rouse, W.B., & Morris, N.M. (1986). On looking into the black box: Prospects and limits on the search for mental models. *Psychological Bulletin*, *100*, 349–363.

Rumelhart, D.E. (1977). Understanding and summarizing brief stories. In D. LaBerge & S.J. Samuels (Eds.) *Basic processes in reading: Perception and comprehension* (pp. 265–303). Hillsdale, NJ: Erlbaum.

Rumelhart, D.E., & McClelland, J.L. (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Schank, R.C. (1975). The structure of episodes in memory. In D.G. Bobrow & A.M. Collins (Eds.), *Representation and understanding studies in cognitive science* (pp. 237–272). New York: Academic Press.

Schank, R.C., Collins, G.C., & Hunter, L.E. (1986). Transcending inductive category formation in learning. *Behavioral and Brain Sciences*, *9*, 639–686.

Smith, E.E., Langston, C., & Nisbett, R.E. (1992). The case for rules in reasoning. *Cognitive Science*, *16*, 1–40.

Smith, V.L. (1991). Prototypes in the courtroom: Lay representations of legal concepts. *Journal of Personality and Social Psychology*, *61*, 857–872.

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. 53–120). Norwood, NJ: Ablex.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*, 435–502.

Thagard, P. (1991). *Probabilistic networks and explanatory coherence*. Unpublished manuscript, Cognitive Science Laboratory, Princeton University.

Trabasso, T., & Sperry, L.L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language*, *24*, 612–630.

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., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315.

van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Van Wallendael, L.B. (1989). The quest for limits on noncomplementarity in opinion revision. *Organizational Behavior and Human Decision Processes*, *43*, 385–405.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, N.J.: Princeton University Press.

Voss, J.F., Perkins, D.N., & Segal, J.W. (Eds.) (1991). *Informal reasoning and education*. Hillsdale, NJ: Erlbaum.

Wilensky, R. (1983). *Planning and understanding: A computational approach to human reasoning*. Reading, MA: Addison-Wesley.

Wolfe, M., & Pennington, N. (1983). *Memory and judgment: Availability versus explanation-based accounts*. Unpublished manuscript, Psychology Department, University of Colorado, Boulder.