A Computer Simulation Model of Juror Decision Making

Daniel M. Gaines*
Computer Science Department
WPI, Worcester, MA 01609
dmgaines@cs.uiuc.edu

David C. Brown
Computer Science Department
WPI, Worcester, MA 01609
dcb@cs.wpi.edu

James K. Doyle
Social Science Department
WPI, Worcester, MA 01609
doyle@wpi.wpi.edu

January 25, 1995

Abstract

Pennington and Hastie's cognitive model of the juror's decision making process, the Story Model, reflects recent attitudes of many cognitive scientists who see an increasing role for stories in human cognition. A computer model of their theory was developed through the use of artificial intelligence techniques. During the design process, the original theory was refined and expanded. The resulting simulation reproduces some empirical evidence for the Story Model and provides support for the importance of stories in decision making.

1 Introduction

This paper describes a computer model of the cognitive processes involved in a juror's decision making task. The computer model is based on the Story Model, a cognitive theory developed and empirically investigated by Nancy Pennington and Reid Hastie of the University of Colorado at Boulder [Pennington and Hastie, 1986] [Pennington and Hastie, 1988] [Pennington and Hastie, 1991] [Pennington and Hastie, 1992] [Pennington and Hastie, unpublished]. In order to determine a verdict, they propose, a juror will construct a story to explain the trial evidence. This story will help the juror to determine the goals of the characters in the trial and reach a verdict.

This model is partially based on the theories of Roger Schank. According to Schank, "Story-telling and understanding are functionally the same thing" [Schank, 1990, p. 24]. He believes that in order to

*Daniel Gaines is currently at: Computer Science Department, UIUC, Urbana, IL 61801, dmgaines@cs.uiuc.edu

understand a story "you have to explain to yourself exactly why the people you are hearing about or observing are doing what they're doing" [Schank, 1990, pp. 6-7]. This implies that by constructing a story to include trial evidence, the juror can be said to have understood the actions of the people involved.

Studying this theory with a computer model offers several benefits. The computer model acts as a *theoretical debugger*, allowing the scientist to observe the theory in action. Specifically, scientists studying the Story Model may be interested in issues such as attention, memory and the interactions between competing stories. The act of expressing a theory in a programming language forces one to be explicit and clear about details. This helps to refine aspects of the Story Model such as how the certainty and acceptance of a story are determined. The Story Model presents a context in which various representations for stories can be explored, improving our understanding of the related artificial intelligence techniques.

The project was broken into three stages. First, Pennington and Hastie's model was studied. A design was developed taking into account the concepts found in the Story Model. The design was implemented in the LISP programming language during the second stage of the project. Finally, several experiments were run with the system to determine its value as a model of Pennington and Hastie's theory and to evaluate the extensions made to the original model.

A brief overview of Pennington and Hastie's Story Model is presented in the next section. Section 3 describes the computer model developed to simulate this theory. The last part of the project, the experimentation phase, is presented in Section 4. Section 5 contains conclusions that can be drawn from this work, as well as suggestions for future research in this area.

2 The Story Model

Pennington and Hastie's theory of the juror's decision making process is called the Story Model because they believe that story construction is the central cognitive process in the juror's task. The Story Model consists of three stages. In the first stage, story construction, the juror constructs stories and selects the one that best explains the evidence. In the next stage the juror learns the verdict categories from the judge. Finally, the juror maps the winning story onto the verdict categories in order to choose a verdict. Most of Pennington and Hastie's research has focused on the first stage of the theory. The research described in this paper concentrates on the first stage exclusively.

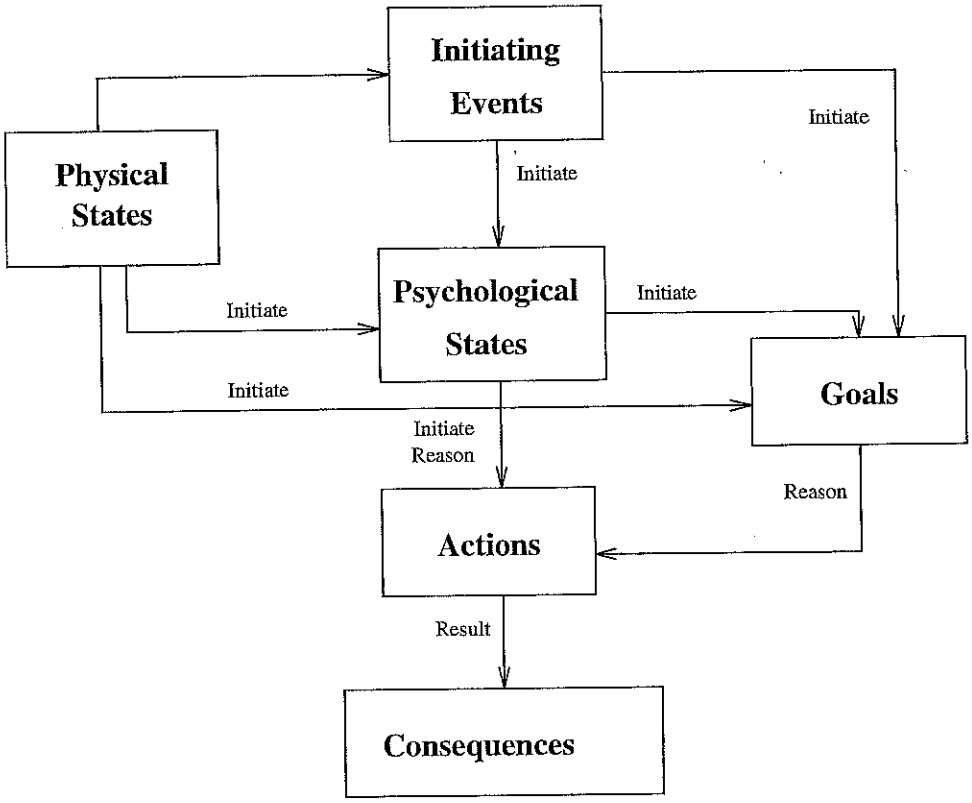


Figure 1: Abstract episode schema [Pennington and Hastie, 1991].

2.1 Story Construction

A juror brings three types of knowledge to bear when faced with a complex decision making task. The juror receives case-specific information during the course of the trial. The jurors also make use of their own general knowledge about how things happen in the world. Finally, the juror uses generic expectations about the structure and content of stories [Pennington and Hastie, 1991].

This third set of knowledge represents the juror's notion of what a story is, how it should be structured and what the importance of its parts are. The mental representation for stories is often described as an episodic schema. Mandler, a psychologist who has performed research in the area of story grammars and schema, describes an episodic schema as, "a mental structure consisting of sets of expectations about the way in which stories proceed" [Mandler, 1984, p. 18]. One such structure is seen in Figure 1, which describes a generic example of an episode used by Pennington and Hastie [1991].

An episode contains role fulfilling events connected by causal relationships. The various components of an episode are *initiating events*, causing characters to have a psychological response and formulate goals; *actions*, performed in a desire to achieve the goals; and *consequences*, states resulting from actions. Because

any event in an episode can itself be an episode, a story can have a hierarchical structure, with the highest level capturing the main theme of the story. With the knowledge of a story's structure the juror is able to form an opinion of a story's completeness, while a story's hierarchical nature helps the juror to identify the most important events [Pennington and Hastie, 1988].

The value of stories in the juror's task is reflected in the following quotation:

> Jurors engage in an active, constructive, comprehension process in which evidence is organized, elaborated and interpreted by them during the course of the trial [Pennington and Hastie, 1991].

Story construction is active, meaning the juror constructs a story during the course of the trial in order to choose a verdict rather than after a trial as an explanation of the chosen verdict. Events in a trial are often not presented in the order in which they occurred. With the aid of stories, jurors are able to impose a causal and temporal ordering on the evidence presented during a trial. The evidence in a trial is often incomplete. With the episodic schema, the juror can detect when information is missing and must be inferred. Finally, information in a trial is often not interesting on its own, but only when related to other pieces of evidence. By making explicit the causal relations between events and grouping together related information, stories help the juror to consider evidence as it relates to previous information gained during the trial.

2.2 Certainty

During the course of a trial the juror may construct several different stories based on the evidence presented. The juror will usually select one of these stories as the best explanation of the evidence and will assign a level of confidence to it, its certainty. Pennington and Hastie state three factors, or certainty principles, that guide the juror in determining this level of confidence; *coverage*, *coherence* and *uniqueness*. The first two principles affect the story's acceptance. A story with high coverage implies that much of the evidence presented in the trial is contained in the story. A story with high coverage will raise the juror's certainty in it and increase the likelihood that it will be accepted as the best explanation [Pennington and Hastie, 1991].

The second principle governing acceptance is coherence. Coherence is determined by *consistency*, no internal contradictions; *plausibility*, events in the story correspond to the juror's beliefs about the way

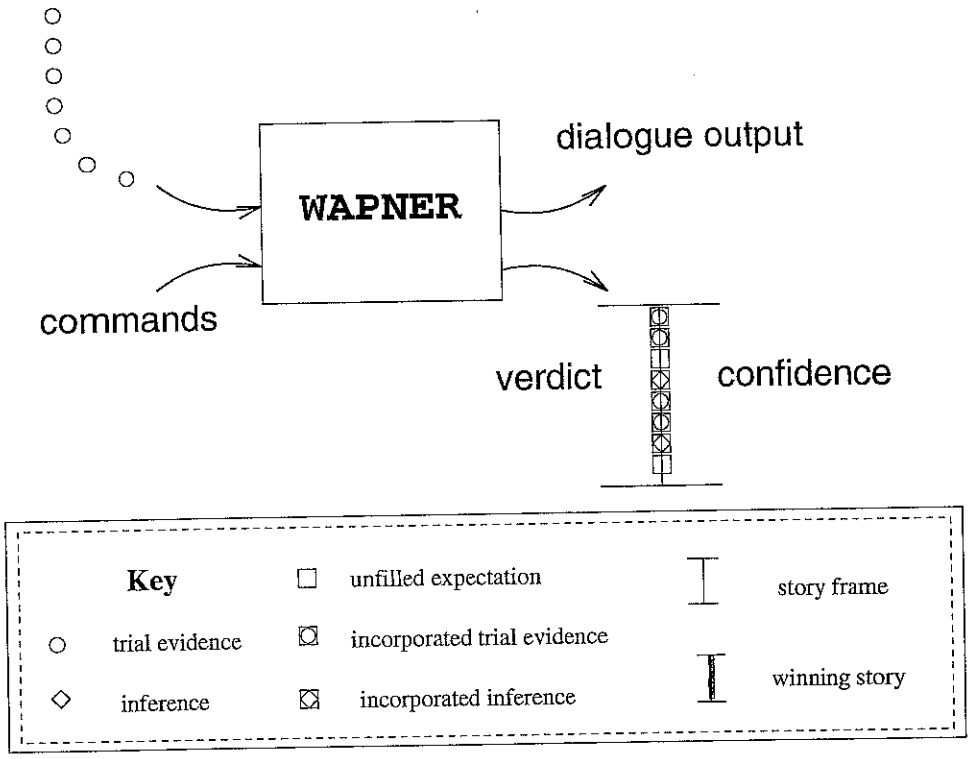


Figure 2: Black box diagram of **WAPNER**.

things happen in the world; and *completeness*, there are no gaps in the story structure. These three components are combined to determine the story's overall coherence.

The third certainty principle, uniqueness, affects the juror's confidence in a given story. The stories will lack uniqueness if more than one of the possible stories is found to be coherent by the juror. In this case the juror will be less confident in them. However, if only one story is coherent the juror will be more confident that that story is the best explanation of what happened [Pennington and Hastie, 1991].

3 WAPNER

WAPNER is a computer simulation model of Pennington and Hastie's Story Model [Pennington and Hastie, 1988] [Pennington and Hastie 1991]. A run of **WAPNER** simulates the juror's decision making process. Figure 2 displays a high level view of the system. The set of trial evidence is presented to the system. During the simulation, **WAPNER** keeps the user informed of its activity through the dialogue output and graphs indicating the stories under consideration. When all of the evidence has been processed, the story that offers the best explanation for the evidence is chosen. Along with the winning story, **WAPNER** outputs a verdict and the confidence it has in that verdict.

3.1 Cognitive Structures

When creating a computer model of a cognitive process, it is essential to adequately represent the cognitive structures involved.

3.1.1 Stories

Stories are the key cognitive structure in the system. **WAPNER** has a set of stories with which it is familiar. These stories correspond to the general world knowledge that the jurors make use of when evaluating the evidence. During the simulation **WAPNER** chooses from this set in order to find a story that best explains the evidence.

Roger Schank, in his work to discover how "people organize the knowledge they must have in order to understand", has developed the idea of scripts [Schank and Abelson, 1977]. Scripts are conceptual units that describe a causally related sequence of events with which we have become familiar. For example, Schank proposes a restaurant script which describes the familiar sequence of events that take place when one enters a restaurant. Depending on the type of restaurant, we might be shown to our seats by a waiter, presented with menus, order our meal, etc. By storing this kind of knowledge in scripts, less processing is required when confronted with familiar situations [Schank and Abelson, 1977].

A variation of scripts was chosen as the representation for stories. There were several reasons for choosing scripts. The main reason is that scripts describe situations that are familiar to people. A story is a sequence of events ordered in a temporal and causal way, which is precisely what scripts are intended to represent. By implementing stories as scripts a particular situation is generalized into a structure that can be applied to similar situations. In this way the juror's ability to relate new information to previous knowledge is reproduced.

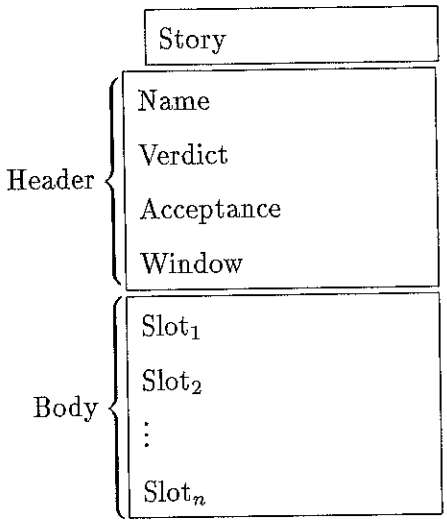


Figure 3: Story structure.

Figure 3 contains a diagram of the story structure used by **WAPNER**. Each story contains a header and a body. The story header includes the name of the story, the corresponding verdict, the acceptance of the story and the story window. Because only the first stage in the Story Model is being simulated, the story construction stage, the verdict category is explicitly attached to the story. Choosing a verdict would normally be done in the third stage of Pennington and Hastie's theory when the juror maps the winning story into one of the categories. **WAPNER**'s opinion of the acceptability of a story is kept in the story header. This acceptance is updated whenever information is added to the story. The story window will be discussed later. Basically, it marks **WAPNER**'s area of focus in the story.

While inspired by Schank's theories, this representation of a story is somewhat simpler than scripts. After our initial experimentations with the Story Model we decided that we did not need the full representational power of a script.

3.1.2 Story Slots

The body of a story is a list of story slots. Figure 4 show the form of a story slot.

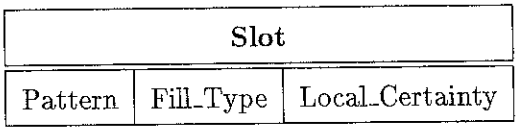


Figure 4: Story slot.

The pattern describes a generalized event. It may contain any number of variables that will be bound with values as indicated by actual evidence. The Fill_Type specifies the state of the slot. A slot may be empty, filled with trial evidence or filled with an inference. **WAPNER**'s confidence in the information filling a particular slot is indicated by the Local_Certainty.

3.1.3 Story Windows

The events in the body are listed in temporal and causal order as would be expected in a story. **WAPNER** will be expecting incoming events to match the ordering in the body. If this does not happen **WAPNER** may not be able to integrate the evidence into the story. Story windows are used to capture this idea. Each story has its own window which is defined by a front pointer and a rear pointer. Each points at a slot in the story. Together they represent the range of evidence **WAPNER** expects to see at any point in time while "listening" to evidence.

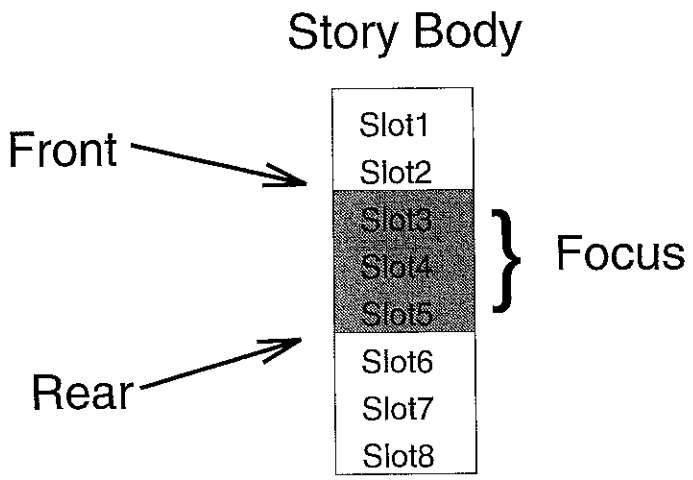


Figure 5: Story window.

A story window, illustrated in Figure 5, is dynamic and will shift during the processing of evidence as **WAPNER**'s expectations for what should come next change. The way in which the window moves can be modified. For the initial design it was decided to leave the front pointer stationary, positioned at the first slot in the story body, and allow the rear pointer to move forward as evidence was filled inside the window. This configuration seems similar to human behavior. When listening to a story it is difficult to make large jumps ahead without notice. Thus the rear pointer is not moved forward very far each time information is added to the story. However, it is fairly easy to insert information into previous portions of a story. By leaving the front pointer at the beginning of the story body, **WAPNER** will have no trouble

adding information into earlier portions.

An alternative would be to allow the rear pointer to move forward along with the front pointer, keeping the window a constant size. Such a design would emphasize the expected ordering of the story because it would be difficult to incorporate evidence that did not follow the order specified by the scripts, as the slots corresponding to that evidence would be outside the window. The size of the window can be altered as well. This is the subject of the third experiment in Section 4.

3.1.4 Active Story

Pennington and Hastie have shown that jurors may be working on multiple stories during their decision making process. Because humans are unable to handle large amounts of information at once, it is unrealistic to allow the system to consider every story at once. For this reason the system chooses an **active story** which becomes the focus of attention. Once a story has been designated as "active", it will be the first story considered when new evidence is encountered. One can think of the active story as what the juror has in mind at a particular moment.

Attention may shift during evidence processing. When a juror hears something that does not match with the active story but is reminiscent of another story the juror may switch attention to a new story. This is called **story swapping**. In this way all of the stories could potentially receive trial evidence even though they are not all active at the same time.

3.1.5 Evidence

The structure of the evidence presented to **WAPNER**, as seen in is similar to that of a story slot, Figure 6. The pattern states the event that this piece of evidence represents. Variables may be included to indicate unknown or unimportant parameters to the event. For example, a witness may know that the defendant had a weapon but may not know exactly what it was.

Evidence		
Pattern	Evidence_Type	Certainty

Figure 6: Evidence.

Evidence may be obtained from the trial or from an inference. The Evidence_Type indicates the origin of

the piece of evidence.

Similar to the Local_Certainty of a story slot, the Certainty of evidence indicates the juror's belief in a given piece of evidence. Because there is no feasible method for correctly determining the credibility of a witness, trial evidence is taken as fact, a certainty of 1. Inferences, however, will be a factor of **WAPNER**'s belief in the world knowledge used to make the inference and the data upon which the world knowledge was applied.

```
((JOHNSON AT GLEASONS AT 21) E 1)
```

Figure 7: Sample evidence.

An example of an evidence can be seen in Figure 7. The evidence item states that Johnson was seen at Gleason's at 9:00 p.m. The E indicates that this is trial evidence. The certainty of the evidence is 1.

3.1.6 Memory

There are two ways in which the system remembers information. When evidence or inferences are incorporated into a story they are also, in a sense, committed to memory. Evidence that finds its way into a story will not be lost.

Because people do not instantly forget things that are not incorporated into a story there is a temporary repository for evidence. If a given piece of evidence does not fit into the active story, the evidence is placed in this memory. Figure 8 contains a representation of the system's memory.

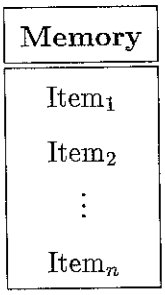


Figure 8: Memory.

This repository has a limited size, and new evidence may overwrite old evidence. While the size of the memory can be adjusted, the original system was constructed with a memory of 5 slots. This is in the range of 7 $\pm$2, the number of slots believed to be in human working memory. The low end of this range was selected in order to penalize information that is not incorporated into a story.

Information sharing is facilitated by the system's memory. When stories are swapped, they get a chance to look into the temporary memory as well as the active story to see if there is any information they can use. Thus stories are allowed to acquire evidence from each other.

3.1.7 World Knowledge

WAPNER's knowledge about the world is kept in the form of rules. Each rule is of the form:

```
(rule-name
  text description
  left hand side
  right hand side
  rule strength)
```

Figure 9 shows an example of one of **WAPNER**'s rules. The rule-name provides a means of identifying the rule. The text description is an English translation of the rule. This is used in the system's output to let the user know what rule is being used.

The left hand side of the rule is a list of preconditions that must be satisfied in order for the rule to be triggered. If any precondition is false then the rule is not applied. Each precondition is a pattern which may contain any number of variables, specified in the form:

```
(? <variable-name>)
```

These variables will be bound to values when matched against the slots of a story, which will serve as a database when attempting to satisfy the left hand side of a rule. Bindings made to variables in one precondition of a rule's left hand side apply to every occurrence of that variable in the rule.

The truth or falsehood of a rule's precondition is determined in two ways. This is most often done by matching the patterns in the left hand side of a rule against filled slots in a story. A hook into LISP was also provided in order to make more powerful associations between variables in a precondition. Thus, rules are allowed to use any functions and procedures that are defined in the LISP environment. Arguments to these functions may include variables which will be bound to their appropriate values before the function is evaluated.

The right hand side of a rule is the conclusion that can be made if the preconditions are met. Normally the conclusion will contain variables referenced in the left hand side. These variables will be instantiated with the values obtained when testing the preconditions.

```
(MURDER-TIME
  "If a person is at the scene within the time
   of murder, that person may be the murderer."
   (((?  PERSON1) AT (?  LOCATION) TIME (?  TIME1))
    ((?  PERSON2) AT (?  LOCATION) TIME (?  TIME2))
    ((?  PERSON1) HAS WEAPON (?  WEAPON))
    ((?  PERSON2) IS DEAD)
    (LISP-CALL (NOT (EQ? '(?  PERSON1) '(?  PERSON2))))
    (LISP-CALL (<= (ABS (- (?  TIME1) (?  TIME2))) 1)))
   ((?  PERSON1) USE (?  WEAPON) ON (?  PERSON2))
   0.6)
```

Figure 9: A sample rule.

The rule-strength ranges from 0 to 1 and defines **WAPNER**'s belief in a rule. The certainty of a rule is determined by the certainty of the left hand side and the strength of the rule. The certainty of the left hand side of a rule is the minimum of the local_certainties of the slots used to satisfy each precondition. A minimizing function is used because a rule is only as strong as its weakest precondition. This value is multiplied by the rule-strength to give the overall certainty of the inference.

In Figure 9, the name of the rule is `MURDER-TIME`. The rule states that if `PERSON1` and `PERSON2` were seen at the same location within an hour of each other, `PERSON1` has a weapon and `PERSON2` ends up dead, one may conclude that `PERSON1` could have murdered `PERSON2`. By using the same variable name to identify `LOCATION` the same location must be used to satisfy both preconditions. This rule demonstrates the use of `LISP-CALL`. This allows rules to access functions defined in LISP. `LISP-CALL` is used in this rule to check the difference in times and make sure that the murderer and the deceased are not the same[1].

Upon successful completion of this rule, the result will be to create a piece of evidence stating that `PERSON1` used a weapon on `PERSON2`. The certainty of this piece of evidence will be the strength of the rule, 0.6 in this example, times the minimum certainty of the slots used to satisfy the preconditions.

3.2 Acceptance and Certainty

A set of equations for calculating the acceptance and certainty of a story were added as refinements to the Story Model. The certainty of a story, according to Pennington and Hastie, is determined by three factors; coverage, coherence and uniqueness. The first two factors describe a story's acceptance while the third

[1]This precondition had to be added to the system because **WAPNER** was becoming confused about suicide and murder.

contributes to the juror's confidence in that story. Section 2.2 contains Pennington and Hastie's definitions of these terms.

$$CV_S = \frac{E_S}{E_T} \tag{1}$$

The coverage of a story, Equation 1, is related to the percentage of evidence that the story takes account of. Thus a story that contains more pieces of evidence will have a higher coverage than a story containing less evidence. E_T is the total amount of trial evidence the juror has seen [2] and E_S is the amount of trial evidence in the story.

$$CH_S = \frac{E_S + I_S}{N} \tag{2}$$

Equation 2 is used to calculate coherence, the second aspect of a story's acceptance. I_S is the number of inferences in the story and N is the number of slots in the story. While coherence is made up of three parts, plausibility, consistency and completeness, only the third, completeness, is explicitly accounted for in this simulation. The others are implied in the script-based implementation of stories. This is a due to the fact that the stories are a reflection of the juror's belief of how things happen in the world. Thus, by definition, these stories will be structured in a coherent and consistent manner. If incoming evidence does not correspond to what the juror believes to be consistent or coherent then it will not be added to the story.

$$A_S = 0.6 \times CV_S + 0.4 \times CH_S \tag{3}$$

In Equation 3, a weighted arithmetic mean is used to combine the coverage and coherence of a story into its acceptance. This was chosen over other combination techniques, such as square and geometric means which give favor to the maximum or minimum, in order to obtain more of an average of the two factors. It was decided that the story's coverage should receive more weight than coherence. This seems reasonable as a story with more hard fact should be chosen over one with many inferences.

$$U_S = A_S - max(A_0 \ldots A_{S-1}, A_{S+1} \ldots A_M) \tag{4}$$

[2] Because the juror will not know the exact amount of evidence that has been presented, a "batch" count is kept. Every time a certain amount of evidence has been seen, 5 items for instance, the global evidence count is increased by that amount. This gives the juror a rough idea of how much evidence has been seen.

After all of the evidence has been processed a winning story is chosen. This will be the story with the highest acceptance value, A_S. Now the level of confidence in this story is determined. This is the same as the story's uniqueness. Equation 4 describes how it is calculated. M is the total number of stories whith which **WAPNER** is familiar.

$$C_S = U_S \tag{5}$$

As indicated in Equation 5, the confidence of the story is equal to its uniqueness.

3.3 Evidence Processing

During the simulation, the system is in a continuous evidence processing loop. This loop is diagrammed in Figure 10. When the system is presented with a piece of evidence, it performs a check to see if it needs to find a new active story. This may be necessary if this is the first piece of evidence and no stories have been activated, or the acceptance of the active story drops below 0. If a new active story is needed, **WAPNER** will search through its story base to find a story that better explains the evidence [3]. **WAPNER** performs a swap with the old active story and the newly chosen story. The newly activated story is allowed to gather up evidence from the previously active story and the memory.

At this stage **WAPNER** is ready to match the evidence with the active story. If a match is made, **WAPNER** adds the information to the appropriate slot in the story [4]. The active acceptance of the story is updated to reflect the acquisition of additional information. **WAPNER** will attempt to apply its world knowledge to the story to see if any new evidence can be inferred. If so, this inference is added to the pool of evidence and will be processed later. If no match was made, the evidence is added to memory and the active acceptance is decremented.

When **WAPNER** runs out of evidence a winning story is chosen. A confidence is calculated and the verdict corresponding to this story is chosen.

[3] **WAPNER** will take into account all of the evidence it "has in mind" at the time. This includes the evidence in the currently active story and the evidence in the short term memory store.

[4] This will likely involve binding variables in the story slot's pattern with values in the evidence. At this time every occurrence of that variable in the story will be bound with this value.

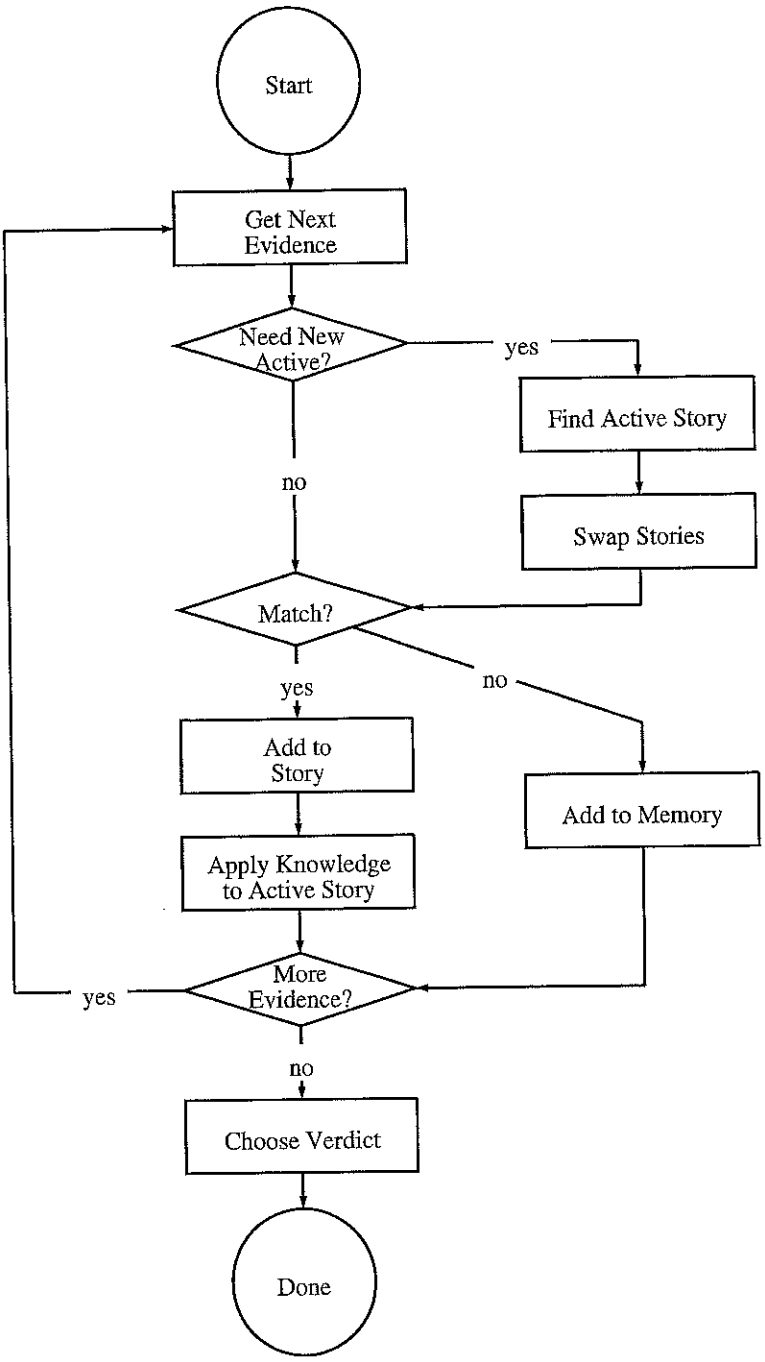


Figure 10: Flow chart of system.

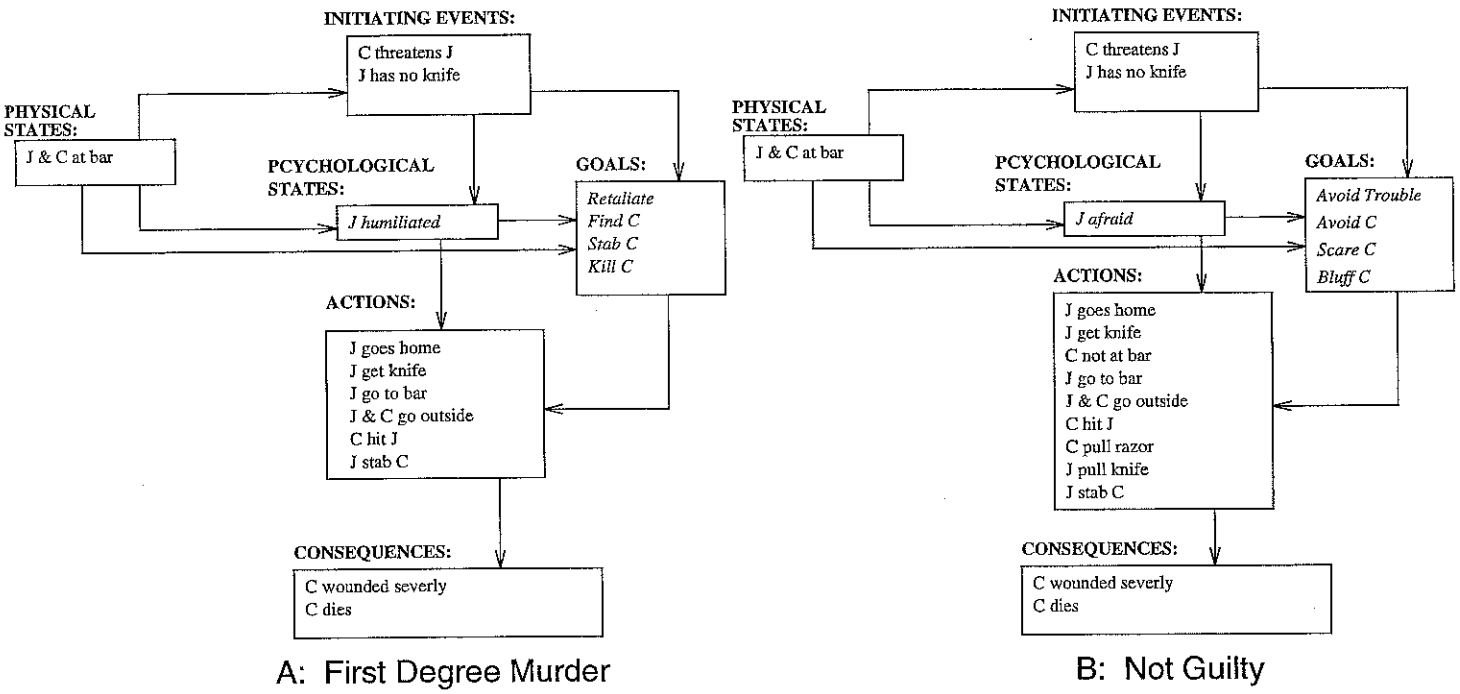


Figure 11: Story structures created by the jurors. The italicized items indicate an inference made by the jurors. (J - Johnson, C - Caldwell) (adapted from [Pennington and Hastie, 1986] [Pennington and Hastie, 1991]).

3.3.1 Using Acceptance

Acceptance is not merely used at the end of evidence processing. Rather, it is an integral part of the juror's decision making. The juror needs a method for determining when it is time to shift attention from the active story and look for another. It seems likely that the juror would use the acceptance of the story as a factor in this decision. Likewise, when the juror goes to choose a new active story, acceptance is the criteria upon which the selection is based.

4 Experiments

Three experiments were performed with the system. The intention was to see if **WAPNER** could reproduce the results obtained by Pennington and Hastie in their experiments with humans. This would demonstrate the validity of the system as a simulation of the Story Model. The experiments were also used to track down any flaws or weaknesses in the system. Finally, these experiments demonstrate the convenience and effectiveness of using computer simulations, such as **WAPNER**, to run experiments.

4.1 Experiment #1 — Evidence Order

The goal of the first experiment was to determine **WAPNER**'s effectiveness in simulating the Story Model. An experiment from Pennington and Hastie's research was presented to the system. If **WAPNER** could produce results similar to their subjects, the system would be deemed a reasonable model of their theory.

Method

This experiment was adapted from Pennington and Hastie's experiment concentrating on the ordering of evidence. They presented an adaptation of the murder trial, the Commonwealth of Massachusetts vs. Johnson. In this case the defendant, Frank Johnson, is charged with the first-degree murder of Alan Caldwell. The events leading to the trial include an argument between Johnson and Caldwell early in the day of Caldwell's death. Caldwell pulled out a razor and threatened Johnson. Later that day, Johnson returned to the bar. Johnson and Caldwell engaged in a fight in which Caldwell was knifed by Johnson. For a more detailed description of the trial please refer to [Pennington and Hastie, 1986].

Taking evidence items from the trial, Pennington and Hastie created two evidence sets, prosecution evidence and defense evidence. Two evidence orderings, story order and witness order, were constructed from each set. The story ordering set listed evidence in the order corresponding to stories created by subjects from an earlier experiment. Figures 11 A and B contain these story structures. Evidence in witness order was listed in the order the evidence was presented during the trial.

Together these orderings comprised 4 evidence sets to be used as stimuli in the experiment. One set contained both prosecution and defense evidence in story order. Another set had prosecution and defense evidence in witness order. The two remaining sets included one set in story order and the other in witness order. Within each set, defense and prosecution evidence were intermixed. All of these sets began with prosecution evidence, as is the case in real trials. One hundred thirty subjects were presented with a tape recorded version of one of the evidence sets. After listening to the tape they were asked to choose a verdict [Pennington and Hastie, 1988].

A similar method was used to construct an experiment for **WAPNER**. The story structures in Figures 11 A and B were translated into script form and became **WAPNER**'s story base. Trial evidence was taken from the trial summary, and codified with respect to **WAPNER**'s representation for stories. Finally, world knowledge was constructed according to the inferences made by Pennington and Hastie's subjects [Pennington and Hastie, unpublished]. Twenty-seven pieces of evidence were taken from the trial. As in

Evidence Set	Order		Verdict	
	Defense	Prosecution	WAPNER	P&H's Subjects
A	Story	Story	G	G (59%)
B	Witness	Witness	G	G (63%)
C	Witness	Story	G	G (78%)
D	Story	Witness	NG	NG (69%)

Table 1: Summary of Experiment #1 results. Forth column contains **WAPNER**'s verdicts. Fifth column lists the verdicts chosen by the majority of Pennington and Hastie's subjects. (Data for Pennington and Hastie's subjects from [Pennington and Hastie, 1988]).

Pennington and Hastie's experiment, the trial evidence was divided into two sets: one with 13 prosecution items, the other with 14 defense items. Two versions of each evidence set were made: one containing evidence in story order, the other in witness order. Combined, this results in the four evidence sets listed in Table 1.

Prosecution and defense evidence is intermixed in the evidence sets, with prosecution evidence listed first. The intermixing of evidence is 1 to 1, that is, one piece of evidence from the prosecution set immediately followed by 1 piece of evidence from the defense set. The effect of other intermixings was studied in experiment #2 (Section 4.2).

Results

The system was run 4 times. Each time a different evidence set was presented as trial evidence. The resulting verdicts and confidences were recorded. Table 1 contains the results found in Pennington and Hastie's subjects and in **WAPNER**.

Figures 12 A-D show graphs of the system during each trial. The graphs represent the state of the system as the simulations progress. After each piece of evidence is presented to the system, the acceptance of each story in the story-base was recorded on the graph. The amount of evidence presented to the system is indicated on the x axis. This can be thought of as a representation of time. As more evidence is seen, more time has progressed in the mock trial. The y axis marks the acceptance of the stories in the story-base. The dotted line labeled "Acceptance Level of Active Story" represents the acceptance of the currently active story. This may be any of the stories in the story-base and will usually switch between stories as

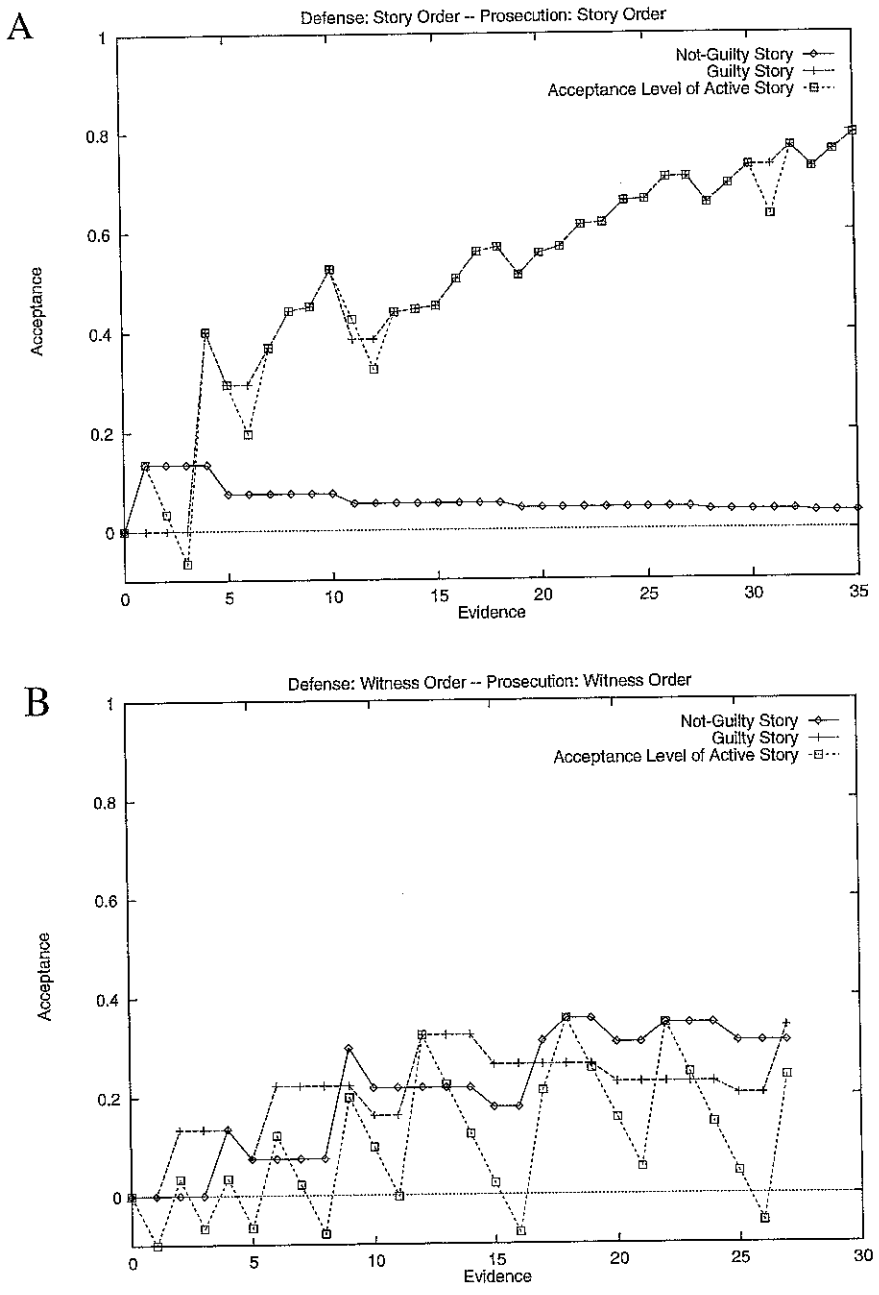
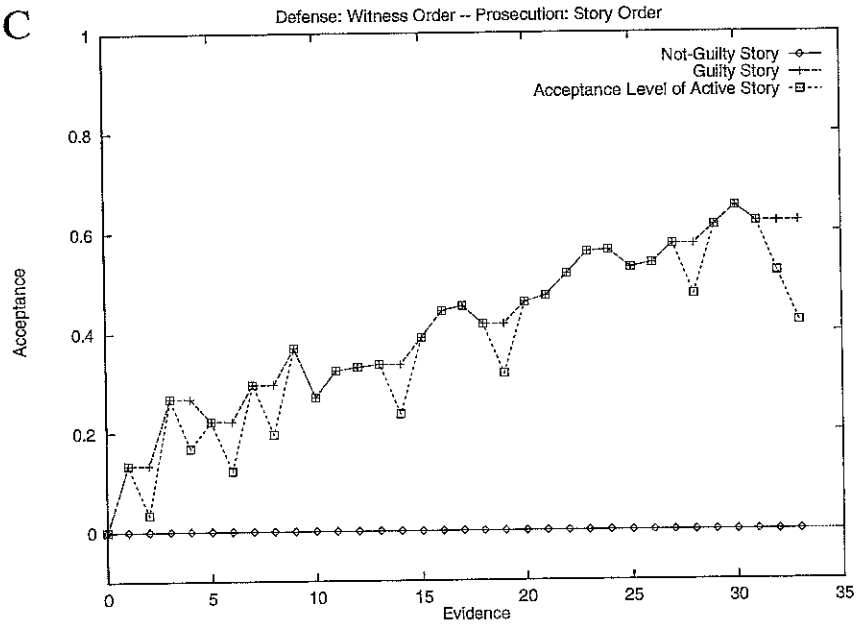


Figure 12 A-B: Results from Experiment #1.

C



D

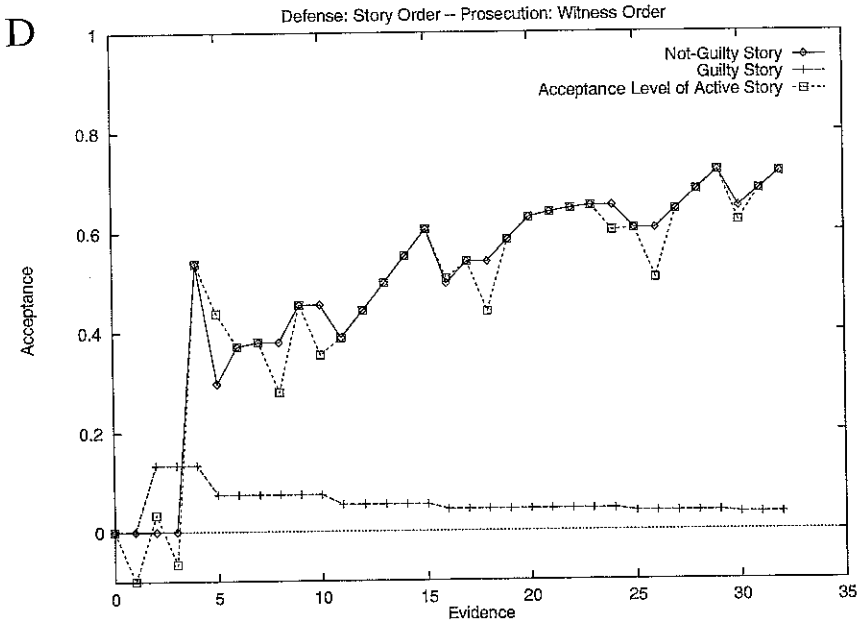


Figure 12 C-D: Results from Experiment #1.

the trial progresses. It is important to note that the "Active" story is not an actual story, but merely indicates which story is currently active in the system (i.e. is the focus of **WAPNER**'s attention). Refer to Section 3.1.4 for a discussion of the active story.

Some details of the trial simulations can be seen in the graphs. Whenever the acceptance of the active story drops below 0, a story swap has taken place. It is easy to tell from the graphs which story is the active story. The acceptance level moves up with whichever story is active. This is due to the fact that only stories which are active can have evidence added to them. The acceptance of stories that have not had evidence added to them tend to drop off during the simulations. This is because one of the factors used to calculate a story's acceptance is inversely proportional to the amount of evidence seen. Thus, a story's acceptance will decrease as it fails to account for more evidence. The verdict chosen by the system is seen in the graphs as well. Whichever story finished with the highest acceptance will be chosen as the winner. This will provide the verdict. The confidence in this verdict is the difference between the winning story's acceptance and the acceptance of the next best story.

The graph in Figure 12A is the result of both prosecution and defense evidence being presented in story order. The not-guilty story became active first, but was soon replaced with the guilty story. The guilty story remained active, missing only a few pieces of evidence, for the remainder of the trial simulation.

In Figure 12B, the system was presented with both prosecution and defense evidence in witness order. The resulting graph represents a series of story swaps. The guilty story just barely wins as the most acceptable story.

Prosecution evidence was presented to the system in story order, with defense evidence in witness order, resulting in the graph of Figure 12C. Here the guilty story wins with the defense story never receiving any evidence.

The opposite presentation was used in Figure 12D, with defense evidence in story order and prosecution evidence in witness order. The guilty story is the first to receive evidence in this graph, but the not-guilty story soon takes over and finishes as the best story.

Discussion

The results were in accordance with the behavior exhibited by Pennington and Hastie's subjects, see Table 1. When **WAPNER**'s results with Pennington and Hastie's subjects, it is important to remember that **WAPNER** is representative of only 1 juror and Pennington and Hastie's results describe a large set

Verdict	Evidence in Story Order							
	Both		Prosecution		Defense		Neither	
	P&H	W	P&H	W	P&H	W	P&H	W
Guilty	3.69	4.80	3.54	4.05	3.36	-	3.30	1.70
Not-Guilty	3.36	-	3.13	-	3.08	4.40	3.08	-

Table 2: Confidence ratings of Pennington and Hastie's subjects and **WAPNER**. Confidence ranges from 1–pure guess to 5–very certain. (P&H = Pennington and Hastie's subjects, W = **WAPNER**.) (Data from Pennington and Hastie's subjects from [Pennington and Hastie, 1988])

of subjects. As Table 1 indicates, **WAPNER**'s verdict is dependent on the ordering of evidence. In the cases where one set of evidence was presented in story order and the other in witness order, such as C and D, **WAPNER** chose a verdict corresponding to the story-ordered evidence set. As the graphs show, the confidences in the decisions were fairly high.

The system was very confident in a verdict of guilt when both evidence sets were presented in story order. This follows Pennington and Hastie's findings where a majority of the subjects chose guilty as a verdict.

The interesting aspect of Figure 12B is **WAPNER**'s relatively low confidence in the verdict, roughly 14%. This indicates that **WAPNER** was unable to construct any stories that clearly explained the evidence. Instead, as the graph shows, **WAPNER** jumped between stories trying to make some sense of the information.

A common pattern in all of the graphs is **WAPNER**'s tendency to swap stories early in the simulation and avoid story swapping as the trial progressed. While nothing definite can be said about human jurors, this does seems consistent with human problem solving in general. Most people are comfortable with changes early on in a task when the cost for such action is relatively cheap. Once some commitment has been placed in a particular decision, however, people tend to stay with it.

It is also interesting to compare **WAPNER**'s verdict confidences with those of Pennington and Hastie's subjects. In Table 2 **WAPNER**'s confidence has been translated into the scale used by Pennington and Hastie: 1–pure guess to 5–very certain. **WAPNER** reproduces the "clarity" factor reported by Pennington and Hastie. That is, subjects were more confident in their verdicts when all evidence was presented in story order giving them a "clear view of the competing interpretations of evidence" [Pennington and Hastie, 1988].

Examining the stories **WAPNER** created provides another explanation for this result. Because the stories in **WAPNER**'s story-base were similar in structure, evidence from either set may be incorporated into both stories. When evidence is presented out of story order, **WAPNER** becomes "confused". Variables are bound incorrectly and evidence is not used and potentially forgotten. Because much of the evidence in each set relates to both stories, both are affected. Thus, if evidence is presented out of order from one set it could hinder story construction in either story. This interaction, which we have referred to as *story interference*, decreases the story's acceptance and, thus, its confidence if it is eventually chosen as the winning story. This suggests that humans may have similar difficulties when constructing competing stories that share a common structure. This suggests that if jurors construct multiple stories that are similar, evidence presented out of story order may adversely affect both stories, leading to a lower confidence in the chosen verdict.

When Pennington and Hastie analyzed the nature of their subject's statements they reported that 55% of their stories were made up of evidence. The remaining 45% was obtained through inferences. They also noted that only 13% of the total trial evidence was incorporated into the subject's stories. Similar results are found in **WAPNER**'s stories. For example, an average story constructed by **WAPNER** is 48% filled with evidence. The remaining 52% is made up of 13% explicit inferences, the direct result of using world knowledge, and 39% assumptions. An assumption means that any unfilled events in the winning story are assumed to have occurred. This is a weak form of an inference relating the trial evidence to the way **WAPNER** expects things to happen in the world. The fact that **WAPNER** reproduces these results is not surprising. The stories the system uses as prototypes were adapted from the stories constructed by Pennington and Hastie's subjects so they share a common structure.

WAPNER used a larger percentage of the trial evidence when constructing stories, 51% as opposed to the 13% usage observed in Pennington and Hastie's subjects. This is due to the fact that a smaller subset of the trial evidence was used in the computer simulation of the experiment.

4.2 Experiment #2 — Evidence Intermixing

The mixing of the evidence sets in the previous experiment was kept constant at 1 to 1 (one piece of evidence from the prosecution followed by one from the defense). The goal of Experiment #2 was to determine the effect of different degrees of intermixing of evidence.

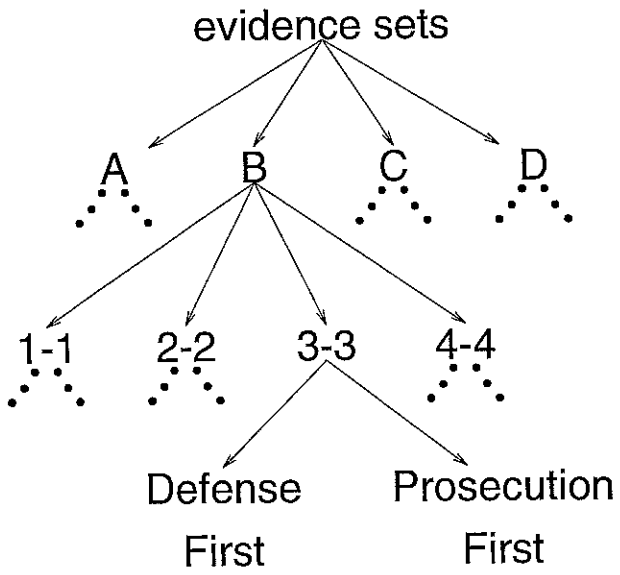


Figure 13: Evidence sets for Experiment #2.

Evidence Set	Previous Verdict	**1–1**	**2–2**	**3–3**	**4–4**
D-SO / P-SO	G	√	√	√	√
D-WO / P-WO	G	×	×	×	×
D-WO / P-SO	G	×	×	√	×
D-SO / P-WO	NG	×	×	√	√

Table 3: Effect of evidence intermixing (√ - affected, × - not affected, D - Defense, P - Prosecution, SO - Story Order, WO - Witness Order).

Method

The four evidence sets from the previous experiment were again used as trial evidence for **WAPNER**. This time, however, the intermixing of evidence was altered within each evidence set. For each set, 4 intermixings were made; 1 to 1, 2 to 2, 3 to 3 and 4 to 4. Each intermixing was presented to **WAPNER** with prosecution evidence first and then with defense evidence first, for a total of 32 distinct evidence sets. Figure 13 illustrates how these conditions break down into the 32 evidence sets.

The verdicts obtained from the previous experiment were used as the control. If **WAPNER** varied from these choices when presented with a different intermixing of one of the original evidence sets, then one can conclude that the intermixing affected the verdict.

Results

Table 3 contains the results from this experiment. The evidence sets correspond to the sets in Table 1. Intermixings that affected the verdict are indicated by a $\surd$. This means that when an evidence set was presented in this order, whichever set of evidence, defense or prosecution, was presented first determined the verdict. Intermixings that had no effect, marked with a $\times$ in the table, resulted in the same verdict as that evidence set had in Experiment #1.

Discussion

Table 3 indicates that in some cases **WAPNER** is overly sensitive to the degree of intermixing of evidence. However, the first row, in which both sets of evidence are in story order, is actually reasonable. When both sets of evidence are in story order, explanations favoring either case are equally easy to construct. Whichever story is begun first will tend to dominate throughout the trial.

It seems likely that a juror would behave similarly. A juror will start constructing a story when presented with evidence. If the juror is presented with two sets of evidence, both of which are in story order, more than one explanation of the evidence could be constructed. In this case, the juror may be swayed by which story was constructed first. This could imply that trials are biased towards the prosecution as evidence from this side is always presented first. However, as Pennington and Hastie have noted, actual trials allow for rebuttals and summaries which lesson the emphasis on evidence intermixing. As this is merely conjecture on our part, this simulation has provided pointers to an interesting series of experiments with human subjects.

The reason for the intermixing effect in the set with prosecution evidence in witness order and defense in story order is a result of the similarity in content of the competing stories. This was mentioned in the discussion section of the previous experiment. Basically, evidence presented out of order from one set has the potential of affecting both of **WAPNER**'s stories. This is more likely when a larger block of witness order evidence is presented at once.

Overall, the results seem positive. Although, no data are available from similar human experiments, **WAPNER**'s behavior when presented with the intermixings seems reasonable. Further experiments should be performed to investigate the affect of constructing similar stories. Further experiments should be performed to investigate the affect that evidence mixing has on humans and correlate them with the results presented here.

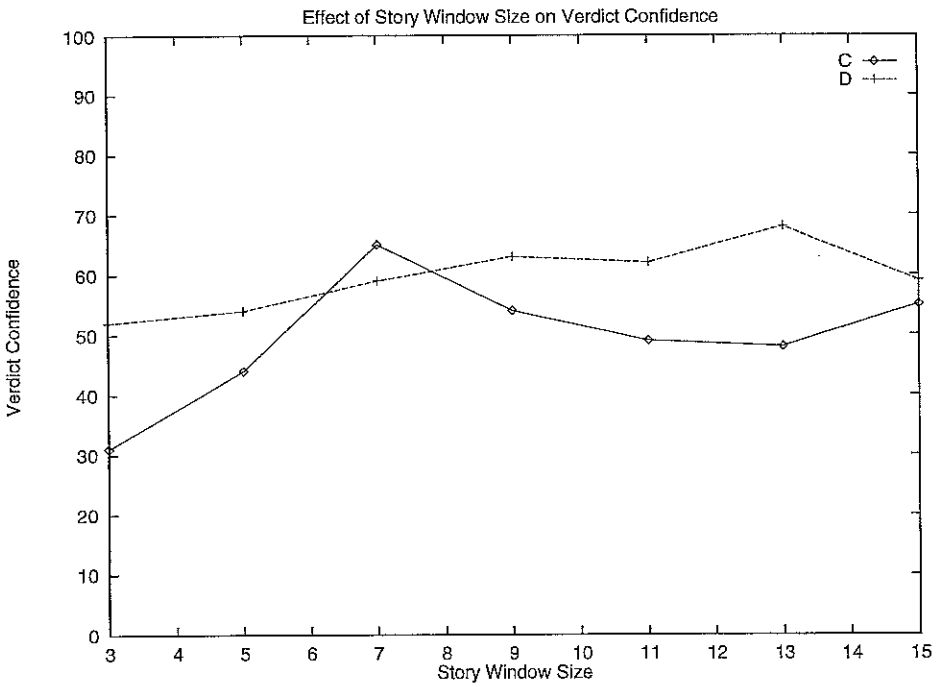


Figure 14: Verdict Confidence vs. Story Window Size.

4.3 Experiment #3 — Story Window Size

WAPNER's point of focus in a story is defined by the story window. Refer to Section 3.1.3 and Figure 5 for a description of the story window. In this experiment the size of the story window is varied to determine the effect on the verdict and verdict confidence.

It was hypothesized that windows of greater size will place less emphasis on evidence ordering. When the window size becomes large, the system's expectation for what events should follow (i.e. what evidence should be heard next) is broadened. Evidence being presented out of temporal and causal order has the possibility of being incorporated into a story that would not normally have accepted the information. With a smaller window size, this story would not be expecting the evidence and would be unable to make use of it. Thus, as the window size increases, obtaining a unique story will become less likely and **WAPNER** will be less confident in the chosen verdict.

Method

Two evidence sets were taken from the first experiment, C, the set with defense evidence in witness order and prosecution evidence in story order, and D, the set with defense in story order and prosecution in witness order. Other evidence sets were not considered as the intent of the experiment was to compare the

affect of the window size on verdict confidence when presented with an evidence set containing both story and witness ordering of evidence. Beginning with a window size of 3 slots, the window was increased by increments of 2 slots until a window of size 15 was reached. The two sets of evidence were presented to the system with each window size.

Results

The verdicts were consistent with the expected results for the evidence set. That is, **WAPNER** chose a verdict of guilty when presented with evidence set C, in which prosecution evidence was listed in story order, and not-guilty when evidence set D was presented, the set in which defense evidence was in story order. Figure 14 plots the level of confidence **WAPNER** had in the verdict against the story window size. For D, the confidence in the verdict grew as the size of the story window grew. However, the verdict confidence began to decline as the window size reached 7 slots. C exhibits this behavior to some extent, but is much flatter than D. The peak for C lies further to the right than the peak for D.

Discussion

The results of this experiment were different from the expected outcome. Increasingly large window sizes *did* cause the confidence level to drop but only after a certain window size had been reached. This seems to indicate that there is some critical window size. Confidence in a verdict grows as the focus increases up to this point.

When the size of the story window is too small, **WAPNER**'s view of the story is impaired to the point that it is unable to make effective use of the story. This is compounded by the fact that some of the information in the evidence set presented in witness order is pertinent to both stories under construction. **WAPNER** is unable to incorporate it into the story because the evidence is not presented in the expected order. Thus, a small window size is too inflexible for practical purposes. One cannot expect every event to follow so strictly to a predefined sequence.

Once the critical window size has been reached, however, the verdict confidence begins to drop off. This is due to the fact that the competing story will become more acceptable as the emphasis on story order is alleviated. Evidence that has not been presented in story order will be incorporated into a competing story if the window size is too large.

The parallel in humans jurors suggests that humans must employ a window size that maximizes the

knowledge contained in the story while minimizing the effect of incorporating information encountered out of sequence. For **WAPNER**, a window size of 9 seems appropriate and, interestingly enough, is consistent with the standard view of the size of human working memory.

C does not demonstrate this behavior as dramatically as D. The slope is less severe and the peak is further along the x axis. This suggests that the story used by **WAPNER** to explain the defense evidence is more "stable", remaining active when evidence is presented in the expected order. This would mean competing stories have little chance of acquiring evidence and decreasing the confidence in the verdict. This stability wears off, however, once the window is opened too wide. The competing story is allowed to collect evidence and have an affect on the verdict.

4.4 General Discussion

The system seems to be a suitable model of Pennington and Hastie's theory. It performed well in the first experiment by successfully reproducing the results obtained in human experiments. **WAPNER** proved convenient to use in creating interesting experiments. The average simulation takes 6 minutes[5]. If enough computers were available, each of the experiments listed above could be performed in less than an hour. The system uncovered new results such as story interference, Experiment #1, and the intermixing effect, Experiment #2, and provided ideas for further experiments with human subjects. These include a closer examination of subjects constructing similar stories and the effect of evidence intermixing.

5 Conclusions

The goal of this work was to translate a cognitive theory of the juror's decision making process into a computer model. During this translation process, it was necessary to expand and refine the original theory. Cognitive issues implicit in the Story Model were made explicit during the design phase. The result was an interpretation of Pennington and Hastie's theory in the form of a computer model.

Expansions and refinements to the system were made with an emphasis on the processes involved in the juror's decision making process. This is different from much of Pennington and Hastie's work which concentrated on the structures used in the juror's task. Some of the expansions to the model include the explicit use of memory and the idea of stories sharing information. The designation of an active story was added along with the story swapping process.

[5]Running Austin Kyoto Common Lisp on a DECstation 5000/240

Refinements were made in the area of the certainty principle. Equations were developed to implement the concepts of acceptance and confidence. These refinements included ways in which the acceptance might be used in other areas of the processing.

Experiments were performed in order to evaluate the system's success as a simulation of the Story Model. Overall the results seem promising. **WAPNER** reproduced the important empirical results observed in Pennington and Hastie's subjects. The system responded in reasonable ways to variations in the evidence sets.

The experiments uncovered some new and interesting results. One such finding was the effect of story similarity. When multiple stories have similar content and structure, as was the case with Pennington and Hastie's subjects, evidence presented out of order can have an adverse effect on all of the stories.

WAPNER provides a convenient means for further development of the Story Model theory. As the experiments have demonstrated, adjustments can be made to the system with ease. The result of the modifications can be obtained quickly. While there is no substitute for experiments with humans, computer simulations, such as **WAPNER**, are an effective tool in developing theoretical models.

The emphasis on this research was on producing a realistic simulation of juror decision making and not directly aimed at exploring the use of stories in decision making in general. However, it is still possible to draw some conclusions about the potential for stories in reasoning tasks. Stories allow one to impose an ordering on large amounts of unordered evidence. As a result, an otherwise senseless mass of disparate information is collected into logically and causally related structures. This structure allows the decision maker to identify where information is missing. An explanation of the evidence is constructed from which a decision can be made. This approach can be applied to domains in which a decision must be made based on a large amount of interrelated information.

5.1 Recommendations for Future Work

There are a number of directions in which the work presented in this paper can be taken. Within the domain of jurors and the Story Model, the stages of the theory not covered in the current system could be developed and integrated into **WAPNER**. Work could be performed to improve aspects of the existing system including a richer representation of stories. It would be nice to have stories constructed more dynamically. A possible approach to this would be the use of story grammars as discussed by Mandler [1984]. Scripts would still fulfill a role in this design by providing prototypes for the stories under construction. However,

the new stories would not have to follow so rigidly to the content of the story scripts.

Further experimentation should be done with the system. Other cases, including civil cases, should be presented to the system. More than two prototypical stories should be introduced to see if the behavior of the system is altered. Several instances of **WAPNER** could be created by providing each with its own unique set of stories and world knowledge. This would allow for the simulation of multiple jurors.

It would be interesting to apply the system to other forms of complex decision making. For instance, Donald Norman has observed the use of stories in business meetings [Norman, 1993]. The design presented in this paper could be applied to this domain. In this task, stories would be used to organize a large amount of information. A business decision would be based upon the best explanation of the information.

5.2 Strengths and Limitations of the System

While the system seems to be a reasonable simulation of the Story Model there are a few limitations. As the major emphasis of this work was on the expansion and refinement of the processes involved in the Story Model, the representation of some of the cognitive structures were not as rich as they might be. This is the case with our representation of stories. A limited version of scripts was used to represent stories. This implementation lacks some of the capabilities of scripts, such as branching and explicit causal relations. A more complex structure is needed to more accurately represent the abstract episodic schema.

There were a few limitations to the experiments. All of the experiments were run with only two prototypical stories, one leading to a not-guilty verdict and the other to a guilty verdict. Actual jurors would make use of many more stories. While the data collected from Pennington and Hastie's research was obtained from many subjects, **WAPNER** represents only one "typical" juror.

It is important to note that despite these limitations, the system has many strengths. **WAPNER** was able to behave in a similar way to Pennington and Hastie's subjects, and the resulting system provides a provocative simulation of the Story Model rich in cognitive theory.

References

[Buchanan et al, 1984] Buchanan, Bruce G. and Shortliffe, Edward H. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Reading, Massachusetts. Addison-Wesley Publishing Company. 1984.

[Collins, 1978] Collins, Alan. "Fragments of a theory of human plausible reasoning." *Theoretical Issues in Natural Language Processing II.* D Waltz (Ed.) Urbana Illinois; University of Illinois Press. pp. 194-201. 1978.

[Collins, 1989] Collins, Alan and Michalski, R. "The logic of plausible reasoning: a core theory." *Cognitive Science. 13.* pp 1-49. 1989.

[Hastie, 1988] Hastie, Reid. "A computer simulation model of person memory." *Journal of Experimental Social Psychology. 24.* pp 423-447. 1988.

[Mandler, 1984] Mandler, Jean Matter. *Stories, Scripts and Scenes: Aspects of Schema Theory.* Hillsdale, New Jersey; Lawrence Erlbaum Associates. 1984.

[Norman, 1993] Norman, Donald A. *Things That Make Us Smart.* Reading, Massachusetts; Addison-Wesley Publishing Company. 1993.

[Pennington and Hastie, 1986] Pennington, Nancy. and Hastie, Reid. "Evidence evaluation in complex decision making." *Journal of Personality and Social Psychology. 51.* pp. 242-258. 1986.

[Pennington and Hastie, 1988] Pennington, Nancy. and Hastie, Reid. "Explanation-based decision making: The effects of memory structure on judgment." *Journal of Experimental Psychology: Learning, Memory and Cognition. 14.* pp. 521-533. 1988.

[Pennington and Hastie, 1991] Pennington, Nancy. and Hastie, Reid. "A cognitive theory of juror decision making: The Story Model." *Cardozo Law Review. 13.* pp. 519-557. 1991.

[Pennington and Hastie, 1992] Pennington, Nancy. and Hastie, Reid. "Explaining the evidence: Tests of the Story Model for jury reasoning." *Journal of Personality and Social Psychology. 62.* pp. 189-206. 1992.

[Pennington and Hastie, unpublished] Pennington, Nancy. and Hastie, Reid. *Reasoning in Explanation-Based Decision Making.*

[Schank and Abelson, 1977] Schank, Roger C. and Abelson, Robert P. *Scripts, Plans, Goals and Understanding.* Hillsdale, New Jersey; Lawrence Erlbaum Associates. 1977.

[Schank, 1990] Schank, Roger C. *Tell Me a Story.* New York, New York; Macmillan Publishing Company. 1990.

[Shortliffe and Buchanan, 1984] Shortliffe, Edward H. and Buchanan, Bruce G. *Rule-based expert systems : the MYCIN experiments of the Stanford Heuristic Programming Project.* Reading, Massachusetts; Addison-Wesley, 1984.