

Narrative Structure, Information Certainty, Emotional Content, and Gender as Factors in a Pseudo Jury Decision-Making Task

James F. Voss and Julie A. Van Dyke
*Learning Research and Development Center
University of Pittsburgh*

Argumentation was studied in a courtroom context in which the prosecuting attorney's summary is assumed to be an argument with "X is guilty" as the claim and the narrative, which contains the evidence of the case, providing support for the claim. In Experiment 1, quality of evidence, narrative coherence, and gender were studied. In Experiments 2A and 2B the role of uncertainty of narrative information, emotional expressions in the narrative, and gender were studied. Both crime-related and non-crime-related uncertain information produced lower guilt ratings and lower ratings of narrative goodness than the baseline, suggesting jury doubt occurs with any narrative uncertainty. Victim-related emotional expressions produced lower guilt ratings than the baseline, although these were mediated by the particular story read. Effects of defendant-related emotional expressions depended on gender and narrative contents. The gender results suggest men respond more heuristically, focusing primarily on evidence, whereas women process the narrative more comprehensively.

Since at least the early Greek period, the courtroom has been an important venue for argumentation. Following a modified version of the dialectic procedure found in Plato's dialogues (van Eemeren, Grootendorst, & Snoeck Henkemans, 1996), it is assumed that questioning individuals who have sworn to tell the truth will lead to the correct knowledge of the events in a particular dispute, as well as to how, why, and by what agency the events occurred. Moreover, in today's courtroom, a defense or prosecuting attorney may deliver an opening or closing statement, with such statements frequently taking the form of narrative accounts of the available evidence. Such statements are structured as arguments with a prosecuting attor-

ney's claim being "X is guilty," and the narrative, which includes a statement of the available evidence, being the support of the claim. This article, which follows from the work of Voss, Wiley, and Sandak (1999), is concerned with factors that influence the processing of a prosecuting attorney's hypothetical narrative in a pseudo jury decision-making task.

How narratives are processed has been the subject of considerable research, including studies of narrative structure (e.g., Mandler & Johnson, 1977), narrative use (e.g., Foucault, 1969/1972), and narrative comprehension (e.g., Trabasso, van den Broek, & Suh, 1989). That the narrative is important to jury decision making was established by Pennington and Hastie (1993), who showed that jurors use trial evidence to construct a narrative encompassing most of the case's important events. The narrative then is related to one of the possible verdicts the judge has explained. Our focus here, however, is not on the narratives constructed during jury deliberation, but on the narrative as provided by a prosecuting attorney's summary statement.

In these studies, after reading a hypothetical prosecuting attorney's statement, participants made judgments of the guilt of the defendant and of qualitative aspects of the prosecutor's narrative. These judgments were of the narrative's convincingness, evidence strength, and overall quality. We refer to these measures collectively as narrative goodness ratings.

In Experiment 1 we were concerned with how the coherence of the prosecutor's statement and the strength of evidence affected guilt and goodness ratings. The study also examined the possible performance differences of participant gender. In Experiments 2A and 2B, we studied the certainty of information and the emotional contents in the prosecutor's statement, as well as the gender of the participants.

NARRATIVE COHERENCE AND EVIDENCE STRENGTH

Experiment 1 is an extension of a previous study (Voss et al., 1999) concerned with whether the quality of a prosecuting attorney's narrative could influence a juror's guilt judgments when narrative quality was varied and the evidence, stated within the narrative, was held constant. In the Voss et al. (1999) experiment, coherence-chronology were studied by comparing guilt and goodness judgments made following a 20-sentence prosecutor's statement, with sentences arranged in a coherent and normal chronological order or in a rearranged order. The rearrangement disrupted the narrative's coherence-chronology but contained the same contents as the normally ordered sentences. The results supported the hypothesis that guilt and goodness judgments would be significantly lower for the poorly ordered narrative, even though the evidence was constant in both types of narrative. The poor narrative thus worked against the prosecuting attorney's goal of achieving a guilt judgment.

Experiment 1 was designed to determine whether this finding would be upheld when coherence–chronology and evidence were both manipulated. Evidence, either strong or weak, and coherence–chronology, either normal or disrupted, were manipulated in a 2×2 design to investigate the role of each variable and their interaction. Of particular interest was whether having a good narrative structure would compensate for poor evidence; that is, whether higher guilt and goodness ratings would be obtained in a poor evidence, highly coherent condition than in a poor evidence, low coherent condition.

Kuhn (2001) recently showed that individual differences occur in juror decision-making situations, which suggests that gender, although it has not received much attention in juror decision-making research, could be a source of individual differences. Specifically, gender of the participant was held orthogonal to the manipulations in Experiment 1 to test a hypothesis derived from Meyers-Levy (1989). In summarizing the literature on gender differences, she suggested that males tend to process information by selecting a single feature, or *cue*, from the situation. The cue is usually regarded to be the most important aspect of the task. This selectivity hypothesis maintains that decision making will be guided primarily by the cue selected. Women, on the other hand, are presumed to be more comprehensive information processors, allowing a number of interrelated factors to influence their decision making. In the application, here, because the task was to determine guilt of the defendant, we assumed that the most relevant feature of the prosecutor's closing statement is the evidence. Thus, the Meyers-Levy hypothesis would predict men would likely use evidence as their cue, and therefore be more sensitive to its manipulation than women, who would be influenced by the broader text features embodied in part by the goodness ratings. Experiment 1 thus was designed to study three issues: the relative importance of evidence and narrative quality, whether a good narrative may compensate for poor evidence, and whether men are more influenced by evidence than women.

INFORMATION UNCERTAINTY

Experiments 2A and 2B also constitute an extension of the Voss et al. (1999) study. Causal relations within a narrative had been determined as one of the characteristics of a good narrative (Leinhardt, Stainton, Virji, & Odoroff, 1994; Pennington & Hastie, 1993). In Voss et al., causal statements were manipulated to test the hypothesis that a prosecuting attorney's narrative with more explicit causal statements would produce higher guilt ratings and goodness judgments than a narrative having less explicit causal statements. An example of this difference is "Matthew's were the only fingerprints on the bat," as opposed to "Matthew's were the only fingerprints on the bat, indicating that he had touched it and, that no one else could have touched it, unless they were wearing gloves."

The results in the original study supported the hypothesis, but some of the changes made in the manipulation were not clearly related to causal factors in-

volution of the crime. For this reason, Experiments 2A and 2B were conducted with a refinement in the manipulation. In Experiment 2A, the variable that we hereafter refer to as information certainty or uncertainty involved two different narratives by the prosecutor. One stated definite information, whereas the other stated uncertain information in which only particular criminal- or crime-related statements were made uncertain. In Experiment 2B the same baseline narrative was employed, but uncertainty was created only in information that was irrelevant to the criminal or crime. Guilt and goodness judgments were obtained in both studies, so it was possible to determine whether, relative to a baseline containing definite statements, crime-related uncertainties produced lower than baseline guilt and goodness judgments and whether crime-irrelevant uncertainties produced similar effects. If the crime-related uncertain information produced lower guilt and goodness ratings than the baseline, but the irrelevant uncertainties did not, then the hypothesized effect of crime-related uncertainty influencing guilt and goodness judgments is supported. However, if the non-crime-related uncertain information also yielded lower guilt and goodness ratings, then the role of uncertain information may be regarded as general.

NARRATIVE AFFECT OR EMOTION

Within the hypothesized courtroom context, a prosecutor's narrative may contain expressions that have affective or emotional contents. Moreover, there are at least two types of such statements, one set pertaining to the defendant and the other to the victim. For example, a prosecutor's argument may include, "He brutally smashed the victim on the back of his head." With respect to the victim, statements such as the "poor and unsuspecting victim" or "the blood-soaked shirt of the victim" may be included. In both cases, these statements are intended to convey a more vivid image of the crime and its severity, either of which may produce higher guilt and goodness ratings than the baseline. Moreover, previous research (Gernsbacher, Hallada, & Robertson, 1998) has shown that readers incorporate emotional information into their situational model of events described in the narrative. We assume therefore that this information becomes part of the representation that may be used by jurors in making their guilt and goodness decisions.

Experiment 2A involved a comparison of performance in two conditions: the baseline, in which the prosecutor's narrative contained no emotional statements, and a crime- or criminal-related condition that contained a number of crime-related (or criminal-related) emotional expressions. In Experiment 2B, there was the same nonemotional baseline condition as in Experiment 2A, but the narrative of the second condition contained emotional statements pertaining to the victim. The issue under study thus was whether emotional statements in the prosecutor's narrative, either crime-related, defendant-related, or victim-related, would act to produce higher guilt and goodness judgments than a baseline narrative void of emotional statements.

PARTICIPANT GENDER

In Experiments 2A and 2B, gender of the participant was orthogonal to the other variables. Following the Meyers-Levy (1989) position, it was hypothesized that if, as previously suggested, men focus more on a specific cue such as evidence, then in Experiment 2A, when the information is crime-related and uncertain, men should provide lower guilt and goodness ratings relative to the baseline than women. However, in Experiment 2B, when the information uncertainty is not crime-related, men should not be influenced by the uncertainty (as compared to baseline performance). Women on the other hand should be influenced by the irrelevant uncertainty because they attend to a broader range of text factors in making their decision, and they would provide lower guilt and goodness ratings than focused in the baseline. Also, with respect to the emotion manipulation, men should not be influenced by either the Experiment 2A or Experiment 2B emotional statements because they are focusing on evidence. However, women in Experiment 2A should produce higher guilt and goodness judgments than men because their processing, assumed to be more comprehensive, would be influenced by the emotional content. In Experiment 2B the same result should occur; that is, there should be no influence of the narrative's emotional content for men, but higher guilt and goodness judgments for the victim-based emotion narrative than the baseline should be obtained for women.

EXPERIMENT 1

Experiment 1 was designed to investigate the possible interaction of the level of coherence–chronology with evidence strength. In addition, gender was held orthogonal to both narrative structure and evidence quality to assess the likelihood that gender differences occur in interpreting the evidence.

Method

Participants. Sixty-four undergraduates (32 men and 32 women) from the University of Pittsburgh psychology participant pool participated in the experiment. The participants received credit as part of their introductory psychology class.

Design and materials. The experiment had a 2×2 factorial design, with the variables evidence (strong, weak) and narrative structure (high coherent–chronology, low coherent–chronology). To allow each person to serve in each of the four conditions, four narratives were constructed, each with four versions. Each narrative version is appropriate to one of the four experimental conditions. Each of the

four narratives described the events of a murder: One was about a man charged with murdering his ex-wife. Another is about a man who is killed because he accidentally hit the defendant's sister while driving home from work. A third story tells of a man who is killed on a hunting trip after the defendant discovers the victim had an affair with his wife. The last narrative is about a pharmacy owner who is killed while his store is being looted. See Appendices A, B, and C for the baseline conditions of the stories not discussed in the text. In all cases, the baseline story was written so that the evidence against the defendant is substantial but not conclusive. This judgment was confirmed by a pilot study. To construct the low-evidence condition, changes were introduced into each of the four baseline stories that admit an alternative explanation of the events and thus shed doubt on the defendant's guilt. For example, the baseline condition text for "The Earthquake" was as follows:

We are all familiar with the earthquake that happened on Tuesday, December 15th. For hours it was pitch black throughout the city. In addition to the power outage, many shops in the downtown area had their windows broken and this led to widespread looting. Because they were concerned about losing their inventory, many shopkeepers stayed in their stores until the police came to secure the area. Vince Morelli was one of these shopkeepers, staying in his pharmacy in order to protect his goods.

The defendant, Edward di Cicco, was looting Vince Morelli's pharmacy when he realized that the owner was present. He continued to loot the store anyway. Edward pointed his gun at Vince Morelli and demanded the keys to the drug cabinet. He was given the keys. Edward then went behind the counter where the drugs were kept, opened the cabinet, and began to put some drugs in a bag he had brought with him. Vince Morelli was not going to let this happen. When Edward was stuffing the drugs into his bag, Vince Morelli reached under the cash register where he kept his gun, but he was not quick enough. Edward di Cicco shot and killed him, and then attempted to flee. As he came out of the pharmacy, two police officers grabbed him. The officers found that the bag he was carrying contained morphine and other drugs. While he was not carrying a gun, they found a gun inside the pharmacy. The only fingerprints found on the gun were those of Edward di Cicco. Edward di Cicco claims that the pharmacist was dead when he entered the store and that he picked up the gun to move it away from the victim when he was attempting to find the keys. The ballistics report confirmed that this was the gun used to kill Vincent Morelli. The defendant should be found guilty.

The low-evidence version of this text contained the changes cited next. The first part of each item is what was in the original baseline text and the second part states the modified contents in the low-evidence text.

1) → There were no eyewitnesses to this crime. (New sentence added); 2) The defendant, Edward di Cicco, was looting Vince Morelli's pharmacy when he realized that the owner was present. → The defendant, Edward di Cicco, was one of several

men looting Vince Morelli's pharmacy when he realized that the owner was present; 3) The only fingerprints found on the gun were those of Edward di Cicco. → No fingerprints were found on this gun, however; 4) Edward di Cicco claims that the pharmacist was dead when he entered the store and that he picked up the gun to move it away from the victim when he was attempting to find the keys. → Edward di Cicco claims that the pharmacist was dead when he entered the store and that another looter killed him; 5) The ballistics report confirmed that this was the gun used to kill Vincent Morelli. → The ballistics report was inconclusive.

Changes 1, 3, and 5 added ambiguity to the account given by the prosecuting attorney. Changes 2 and 4 added alternative suspects.

To create the low-coherence–chronology narratives, we followed the procedure used in Voss et al. (1999), in which the order of sentences from the baseline narrative was randomized. The order of the sentences for the low-coherence version of the earthquake text was 6, 2, 15, 1, 7, 4, 5, 13, 16, 10, 14, 19, 17, 8, 11, 3, 12, 9, 18, 20. To preserve local coherence, minor sentence modifications were made such as replacing anaphors with the appropriate referent name.

These story manipulations were checked in a norming study with a group of 11 participants who were given all versions of all the narratives. The narratives were counterbalanced for condition and story order by using a Greco-Latin square. The participants were asked to judge the defendant's guilt, and the convincingness, quality of the evidence, and overall quality of the prosecutor's narrative. We used the narrative quality and evidence results to verify that our manipulations created the intended effects, those of lower guilt ratings and low-evidence quality with the low-evidence text and lower coherence–chronology ratings with the low-coherence–chronology text. Before conducting Experiment 1, some additional changes were made to the narratives as suggested by these results. We also examined mean ratings for each condition by story to ascertain whether the four stories were of comparable baseline quality and made changes to stories that appeared to deviate from the ratings obtained for the other three.

Procedure. Participants served in groups of 8. Each participant was given a packet containing the four different prosecutor's statements, each representing one of the four experimental conditions. Participants were instructed to assume the role of a juror. The four narratives and four experimental conditions were counterbalanced using a Greco-Latin square. The experimenter explained that each narrative represented the prosecuting attorney's closing argument in a court case and based on this argument, the participants were to make judgments about the guilt of the defendant.

Measures. After reading each narrative, each participant was asked to provide a rating of guilt on a scale from 1 to 11, annotated to maximize rating comparability between participants. For example, a rating of 9 was annotated as "I think X is guilty, but I have some doubt," whereas a rating of 10 was annotated "I am

positive X is guilty, but I have a little doubt.” The extreme ratings were 11 (*I am positive X is guilty*) and 1 (*I am positive X is not guilty*). Additionally, participants were asked three questions regarding the narrative itself. On an annotated scale of 1 (*lowest*) to 5 (*highest*), students were asked to judge convincingness of the prosecutor’s argument, quality of the argument, and strength of the evidence. Annotations for these scales used language such as *not at all* or *rather* to indicate variation at the low end, and *very* and *extremely* to indicate variation at the high end.

Results

Table 1 presents the means for all statistically significant judgment differences of Experiment 1 as well as means for a few nonsignificant differences.

Guilt ratings. Guilt judgments were higher in the strong evidence than in the weak evidence condition, $F(1, 224) = 96.84, p < .001$. However, the mean guilt rating did not vary significantly as a function of narrative condition, $F(1, 224) < 1$. Although there was little difference in mean guilt judgment for the two weak evidence conditions (6.09 vs. 6.19), the mean guilt judgment in the strong evidence and the baseline narrative compared to the mean of the strong evidence and disruptive narrative condition did not produce a significant Evidence \times Narrative interaction, $F(1, 224) = 1.54, p < .21$. The failure of low-evidence conditions to

TABLE 1
Guilt and Goodness Ratings for the Evidence and
Narrative and Selected Interactive Conditions of Experiment 1

Condition	Guilt	Quality	Convincingness	Strength
Evidence (E)				
Strong (S)	8.93***	3.13***	2.97***	3.44***
Weak (W)	6.14	2.15	1.74	1.90
Narrative (N)				
Baseline (B)	7.66	3.16***	2.62***	2.82***
Disrupted (D)	7.41	2.11	2.09	2.52
Evidence \times Narrative				
SB	9.23	3.89**	3.38***	3.73***
SD	8.63	2.36	2.56	3.14
WB	6.09	2.44	1.86	1.91
WD	6.19	1.86	1.63	1.89
Evidence \times Gender				
S Male	9.33*	3.27*	3.20**	3.53
S Female	8.53	2.98	2.73	3.34
W Male	5.94	2.03	1.66	1.78
W Female	6.34	2.27	1.83	2.02

* $p = .05$. ** $p = .01$. *** $p = .001$.

yield higher guilt ratings in the coherent condition than in the disruptive condition does not support the idea that a more coherent narrative will compensate for weak evidence. The goodness ratings do suggest, however, that with strong evidence, a more coherent narrative may produce higher guilt ratings.

Gender did not significantly influence guilt judgments, with mean male and female judgments of 7.63 and 7.44, respectively ($F < 1$). However, the interaction of gender and evidence showed that men made more extreme evidence-based guilt judgments than women, $F(1, 224) = 4.51, p < .04$. In other words, compared to women, men made higher judgments of guilt when evidence was strong and lower judgments of guilt when evidence was weak, a result suggesting that men weighted evidence, or the lack thereof, to a greater extent than women. This result provides some support for the Meyers-Levy (1989) hypothesis that men focus more on a specific cue—in this case, evidence—than women.

A significant story effect was obtained, with the Story 2 mean guilt rating being significantly lower than the means of Stories 1 and 4. Table 2 presents the mean guilt ratings for each of the stories, $F(3, 224) = 6.81, p < .001$. The means of the Story \times Evidence interaction, $F(3, 224) = 4.22, p < .01$, are presented in the columns for strong and weak evidence under guilt ratings. The high evidence ratings for the four stories varied only from 8.69 to 9.25, whereas the low evidence ratings varied from 4.50 to 7.53. This result suggests that the difference in guilt ratings as a function of story and evidence is due to the relative differences in the weak evidence condition, primarily in Story 2. A possible reason for the low guilt ratings in that story is that in the low-evidence condition it is explicitly stated that it was “difficult to see” the victim, increasing the likelihood that the shooting

TABLE 2
Mean Guilt and Goodness Ratings for the Story and the Story \times Evidence Conditions in Experiment 1

Story	Guilt			Quality		
	Baseline	Strong	Weak	Baseline	Strong	Weak
1	8.00***	9.25***	6.75	2.89	3.41**	2.38
2	6.64	8.78	4.50	2.52	3.34	1.69
3	7.23	8.69	5.78	2.55	2.94	2.16
4	8.27	9.00	7.53	2.59	2.81	2.38

Story	Convincingness			Strength		
	Baseline	Strong	Weak	Baseline	Strong	Weak
1	2.70***	3.47**	1.94	2.92**	3.78*	2.06
2	2.17	2.97	1.38	2.42	3.41	1.44
3	2.06	2.50	1.63	2.44	3.03	1.84
4	2.48	2.94	2.03	2.89	3.53	2.25

* $p = .06$. ** $p = .01$. *** $p = .001$.

could be regarded as accidental. Furthermore, the possible motive of the victim having an affair with the defendant's wife was deleted in the low-evidence condition, whereas the original motive remained intact in the low-evidence condition for the other three stories. These two modifications in Story 2 likely combined to make the alternative of an accidental killing plausible, thereby producing a relatively low guilt rating compared with the evidence manipulation changes of Stories 1 and 4. Story 3 also had a substantial good and poor evidence difference, and this story appears to have a feasible alternative hypothesis in the weak condition, as the text suggests that someone other than the defendant may have killed the victim.

Goodness judgments. The quality judgment served as a check on the narrative manipulation. Narrative quality (see Table 1) was judged better in the coherent condition than the disruptive condition, $F(1, 224) = 61.36, p < .001$, and also better in the good evidence compared to the poor evidence condition, $F(1, 224) = 52.61, p < .001$. Also, although having a more coherent narrative produced higher quality ratings for both the strong and weak evidence conditions, the effect was substantially greater when the evidence was strong, $F(1, 224) = 12.53, p < .001$. This same set of results involving the evidence and narrative main effects and their significant Narrative \times Evidence interactions held for judgments of convincingness, $F(1, 224) = 97.54, p < .001$; $F(1, 224) = 17.76, p < .001$; $F(1, 224) = 5.42, p < .02$; and strength, $F(1, 224) = 123.87, p < .001$; $F(1, 224) = 4.86, p < .03$; $F(1, 224) = 4.37, p < .04$, with F s presented for the evidence effect, narrative effect, and Evidence \times Narrative interaction, respectively. These results indicate that the prosecutor's summary was judged better when it had both strong evidence and a more coherent narrative. Furthermore, the goodness rating interactions of evidence and narrative were also consistent with the guilt ratings in that a more coherent narrative was more effective in obtaining a guilty verdict when the evidence is strong compared to when it is weak. However, despite the narrative main effect and the Narrative \times Evidence interaction obtained for the goodness ratings, the differences in narrative structure did not produce a significant effect of narrative or the Narrative \times Evidence interaction in the guilt ratings. This difference suggests that goodness judgments are derived from narrative properties, but that judgments of guilt may be more restricted to evidential characteristics of the narrative.

None of the goodness ratings approached a significant gender effect, but the Gender \times Evidence interaction was significant for convincingness, $F(1, 224) = 6.65, p < .01$, and was of borderline significance for quality, $F(1, 224) = 3.67, p < .06$. For argument strength, $F(1, 224) = 2.33, p < .13$. The means of the Gender \times Evidence interactions, shown in Table 1, indicate that although significance varied for the three goodness ratings, the pattern of the goodness ratings was the same as the pattern of the guilt ratings; that is, men gave more extreme judgments than women in both the strong and weak evidence conditions.

Story effects were significant for two of the three goodness ratings: quality, $F(1, 224) = 1.64, p < .18$; convincingness, $F(1, 224) = 5.55, p < .001$, and strength, $F(1, 224) = 3.97, p < .01$. The means for these three ratings are presented in Table 2. As with the guilt ratings, all three goodness ratings for Story 2 were quite low, and the ratings for Stories 1 and 4 were highest. In the convincingness ratings, the mean of Story 1 differed from that of Stories 2 and 3. The Story \times Evidence interaction was significant for the quality judgments, $F(3, 224) = 3.65, p < .01$; borderline for convincingness, $F(3, 224) = 2.45, p < .06$; and not significant for strength, $F(3, 224) = 1.77, p < .15$.

Discussion

The first issue of Experiment 1, as mentioned earlier, concerns the relative role of evidence and narrative quality. These results indicate that if the evidence is poor, guilt and goodness judgments will be relatively low, regardless of the quality of the narrative. This finding does not support the hypothesis that a good narrative can compensate for poor evidence. With respect to narrative quality, Klettke and Graesser (2000) found that if a narrative contains contradictions, poor guilt ratings result. Although in this study the “jurors” gave higher goodness ratings for the coherent than for the disrupted narrative, narrative quality ratings did not influence guilt ratings. The findings of Voss et al. (1999), in which a disrupted narrative did produce significantly lower guilt and goodness ratings, taken with the present lack of an effect on guilt ratings suggest that the narrative quality may play a role when the evidence is neither good nor poor, but instead is somewhat uncertain. The other result regarding the interaction of evidence and narrative is that a good narrative does enhance goodness ratings when evidence is strong, with the guilt ratings showing the same tendency, although not significantly so.

Results involving the gender variable indicated that men made more extreme guilt and goodness judgments than women for both good evidence and poor evidence conditions. Men apparently gave stronger weight to evidence characteristics than women, or men focused primarily on the evidence information whereas women focused less on one aspect of the narrative (Meyers-Levy, 1989).

The results also show that although care was taken to verify similar ratings for each story on each of the relevant dimensions, story effects nevertheless occurred. As stated, the story effects seem due to more pronounced effects of the weak evidence conditions in Story 2 and to some extent in Story 3 with respect to the availability of alternative hypotheses regarding the defendant’s guilt.

EXPERIMENTS 2A AND 2B

Experiments 2A and 2B, as previously mentioned, focused on three issues. One was the extent to which guilt and goodness ratings are influenced by the relative certainty of the contents of the statement, with such information being crime-

related or non-crime-related. The second issue was the extent to which emotional statements related either negatively to the defendant or crime or positively toward the victim influence guilt and goodness judgments, as compared to a narrative not containing emotional statements. The third issue was the extent to which participant gender influences guilt and goodness judgments and the extent to which interactive effects occur between the certainty and the emotional manipulations and the participant gender. Because there was a need for a within-subjects design, but for practical reasons a design with a limit of four conditions, two 2×2 experiments were simultaneously conducted.

EXPERIMENT 2A

Method

Participants. Ninety-six University of Pittsburgh undergraduates (48 men, 48 women) participated in Experiment 2A for partial course credit.

Design and materials. Experiment 2A had a 2 (certainty) \times 2 (emotion) factorial design, crime-relevant uncertain information versus certain information (baseline) and defendant-related emotional expressions versus no emotional contents (baseline). Thus, there was a baseline condition (certain information and no emotional contents), no emotional contents and uncertain crime-related information, certain crime-related information and defendant-related emotional contents, and a condition with both uncertain crime-related information and defendant-related emotional contents.

The four baseline stories from Experiment 1 were used in Experiment 2A. Taking "The Earthquake," Story 1, presented earlier as a baseline, the following five changes were introduced to create the crime-relevant probability manipulations.

The defendant, Edward di Cicco, was looting . . . → The defendant, Edward di Cicco, was probably looting . . . ; 2) Edward pointed his gun at Vince Morelli . . . → Edward most likely pointed his gun at Vince Morelli . . . ; 3) Edward di Cicco shot and killed him . . . → Edward di Cicco apparently shot and killed him . . . ; 4) The only fingerprints found on the gun were those of Edward di Cicco. → The only fingerprints found on the gun were probably those of Edward di Cicco; 5) The ballistics report confirmed that this was the gun used to kill Vincent Morelli. → The ballistics report confirmed that this could have been the gun used to kill Vincent Morelli.

The criminal-relevant emotion versions were created similarly.

1) He continued to loot the store anyway. → He was so obsessed that he continued to loot the store anyway; 2) Edward . . . → Gesturing wildly, Edward . . . ; 3) Edward

pointed his gun at Vince Morelli . . . → Edward burst into the store, pointed his gun at Vince Morelli . . . ; 4) Edward was stuffing the drugs into his bag . . . → Edward was frantically stuffing the drugs into his bag; 5) Edward di Cicco shot and killed him . . . → Edward di Cicco shot him ruthlessly.

Four text versions for each story were constructed such that there was one version for each condition of the experiment. By using the resulting four versions of each of the four stories, each individual participated once in each of the four experimental conditions and each condition for each participant involved a different narrative, as determined by a 4×4 Greco-Latin square.

A validity check for the manipulations was run using 16 additional participants (9 men and 7 women). Each rated all 16 stories of this study and an additional 12 used in Experiment 2B for argument strength, definiteness, and emotional impact. The stories were presented in blocks, so that the participants saw all the baseline stories first, followed by four additional blocks containing each of the remaining versions. Presentation order was counterbalanced with a Greco-Latin square for story and story version. The uncertainty manipulation resulted in lower ratings for definiteness across all stories except in one case, and lower ratings of story strength were obtained in all but three cases. The emotion manipulations resulted in increased ratings for emotional impact in all but four cases. Subsequently, adjustments were made to the stories in the conditions that did not produce the expected results. We also collected guilt ratings for all stories in the pilot study and made some changes to the baseline versions of individual stories so that the mean guilt rating was similar across all texts.

Procedure and measures. The same procedure described in Experiment 1 was followed here. Participants were asked to provide guilt ratings, followed by quality, convincingness, and—in place of strength ratings—quality of evidence ratings. The same scales described in Experiment 1 were used, with the exception of evidence strength ratings that ranged between 10 (*very strong*) and 1 (*not strong*).

Results

Guilt ratings. Uncertain evidence produced lower guilt ratings at a borderline significance level compared to the baseline condition, $F(1, 352) = 3.51, p < .06$ (see Table 3). Defendant-related negative emotional content did not significantly affect guilt ratings ($F < 1$). Gender also did not influence guilt ratings ($F < 1$).

There was a significant story effect on guilt ratings, $F(3, 352) = 5.34, p < .001$. As suggested by the means in Table 3, guilt ratings on the four stories significantly differed from each other except for Stories 1 and 2. The only significant interaction obtained was that of Emotion \times Story \times Gender, $F(3, 352) = 2.88, p < .04$. The means of this interaction, shown in Table 4, indicate that for all stories,

TABLE 3
Mean Guilt and Goodness Ratings for the Main Effects of Experiment 2A

	<i>Guilt</i>	<i>Quality</i>	<i>Convincingness</i>	<i>Evidence</i>
Information certainty				
Baseline	8.26*	3.67**	3.31***	7.28***
Uncertain	7.75	3.01	2.60	6.35
Emotion				
Baseline	8.03	3.19**	2.80**	6.54**
Emotional	7.98	3.48	3.12	7.09
Gender				
Male	7.88	3.37	3.00	6.67
Female	8.13	3.31	2.92	6.96
Story				
1 (Burglary)	8.01***	3.10***	2.76***	6.52***
2 (Hunting)	8.01	3.43	3.07	6.77
3 (Mob attack)	7.23	3.09	2.64	6.25
4 (Wife killing)	8.77	3.72	3.37	7.72

p* = .06. *p* = .01. ****p* = .001.

TABLE 4
Mean Guilt Ratings of the Story × Emotion × Gender Interaction of Experiment 2A

<i>Story</i>	<i>Male</i>		<i>Female</i>	
	<i>Baseline</i>	<i>Emotion</i>	<i>Baseline</i>	<i>Emotion</i>
1	8.38*	7.13	8.08	8.46
2	8.00	7.63	8.13	8.29
3	6.42	8.25	7.42	6.83
4	9.08	8.17	8.71	9.13

**p* = .05.

the direction of the difference of the baseline and emotion conditions was opposite for men and women. For men, the defendant-related emotion condition produced lower guilt ratings in Stories 1, 2, and 4 and higher guilt ratings in Story 3. For women, however, Stories 1, 2, and 4 produced higher guilt ratings and Story 3 produced lower guilt rating judgments in the defendant-related emotion condition. Examining the stories, men produced lower guilt ratings for stories in which the emotional statement was crime-related, whereas women’s ratings were lower for the story in which the emotional expression was person-related.

In summary, the guilt rating data suggest that making evidence relatively uncertain produced marginally lower guilt ratings, and that neither emotional contents of the crime or gender significantly influenced guilt judgments. There was, however, an interaction of story, gender, and emotion, suggesting that men and women responded differently to the particular emotional contents of each story,

with women responding to emotional statements related to the defendant and men responding to emotional statements pertaining to the crime and evidence.

Goodness judgments. For all three ratings, the pattern of main effects was identical. As presented in Table 3, the means showed that the prosecutor's statement was rated as poorer in the crime-related uncertain condition than in the baseline condition, $F(1, 352) = 7.42, p < .01$; $F(1, 352) = 40.71, p < .001$; and $F(1, 352) = 17.60, p < .001$, for the quality, convincingness, and evidence ratings, respectively. These results are consistent with the guilt rating findings, although the magnitude of the difference in guilt ratings is not as substantial.

As also shown in Table 3, the means of the three goodness judgments were higher in the baseline condition than in the defendant-related emotion condition, $F(1, 352) = 7.42, p < .01$; $F(1, 352) = 7.92, p < .01$; and $F(1, 352) = 6.43, p < .01$, for the quality, convincingness, and evidence ratings, respectively. These judgments indicated that the inclusion of the defendant-based emotion contents produced a judgment of poorer quality, but this effect did not carry over to the guilt judgments. Gender was not significantly related to goodness judgments.

Story had a significant effect on all three goodness judgments. Analyses indicated $F(3, 352) = 8.05, p < .001$; $F(3, 352) = 8.70, p < .001$; and $F(3, 352) = 8.45, p < .001$, for the quality, convincingness, and evidence judgments, respectively. For each rating, Story 4 was rated significantly higher than Stories 1 and 3. This result, moreover, is similar to the guilt ratings previously discussed, as Story 4 ratings were significantly greater than those of the other three conditions, whereas Story 3 yielded ratings lower than those of Stories 1 and 2.

One possibility for these results is that the stories varied with respect to the extent an alternative hypothesis for the crime became more apparent in the crime-related uncertain condition. Story 4, which received relatively high guilt and goodness ratings, did not provide a reasonable alternative to the guilt of the defendant. In both versions, Raymond tells Myrna that he will kill her and his gun was recovered with his fingerprints. The difference of "he presumably shot her" and "he shot her" seems small in the context of an explicit statement that he will kill her and the additional evidence against him. Furthermore, although the evidence against him becomes weak when the uncertain changes are added, there is nothing that would suggest another person may have committed the crime. In Story 3, although the evidence is against Matthew, there was a mob of boys and any one of them could have committed the crime. Hence, in this story, there are lower guilt ratings, possibly because of a more apparent alternative explanation for the events, even in the certainty condition. In Story 1, there is little support for an alternative scenario unless some other looter committed the crime.

A significant Emotion \times Story \times Gender interaction was found, $F(3, 224) = 2.63, p < .05$. This interaction appears to be due to the previously discussed characteristics of Story 3. In this story, men provided much higher evidence ratings in the emotion condition ($M = 7.25$) than in the baseline condition ($M = 5.50$). For

women, the two respective means were 6.00 and 6.25. The men's baseline rating suggests that in the "Hunting Story" men did not consider the evidence for Paul's shooting of Rob strong, perhaps finding a credible alternative explanation of a hunting accident. Although the emotional contents were not related to evidence per se, they did pertain to the description of the crime itself and included statements such as, "Rob lying in the grass, dead, his blood splattered all over the woods," as compared to "Rob lying in the grass, dead," in the baseline condition. The fact that men are more affected by this manipulation than women fits with the Meyers-Levy (1989) selectivity hypothesis, provided it is assumed that the men interpreted the emotional conditions as related to the evidence in the story, an interpretation suggested by the increase in their evidence strength judgments.

With respect to the previously mentioned issues, although crime-related uncertainty produced substantially lower goodness judgments than the baseline, it produced lower guilt ratings only to borderline significance. Furthermore, although negative emotional expressions in the narrative produced higher goodness ratings than those found in the nonemotional condition, guilt ratings were not influenced. These two effects suggest that the prosecutor's narrative is judged as a narrative for its quality, convincing, and evidence contents, but that the guilt judgments are made according to a more restrictive set of criteria. It appears that by taking on the role of a juror, participants judge the narrative in relation to the prosecutor's goal. Consequently, their guilt judgments may reflect additional processing or interpretation based on the jurors' own perceptions of the case.

With respect to gender, the Emotion \times Gender interaction has $F < 1$, but the results indicate that gender may play a role in relation to the particular stories and the particular emotional expressions used in those stories. Specifically, emotion produced higher guilt ratings than the baseline for women for three stories in which the emotional expressions involved the criminals. However, for men, the baseline guilt rating was relatively low, possibly because they thought that a hunting accident was a reasonable interpretation of the crime, but added emotional expressions, which may have affected the quality of the evidence, produced higher ratings.

EXPERIMENT 2B

Experiment 2B, as with Experiment 2A, consisted of manipulations of the certainty of narrative statements and the inclusion of emotional statements. In Experiment 2B, however, the certainty manipulation consisted of the same baseline condition used in Experiment 2A and a second condition in which the textual changes added uncertainty to five aspects of the narrative; however, none of the points was related to the crime. Thus, the question was whether crime-irrelevant uncertainty would influence guilt and goodness judgments. The emotion condition employed the same baseline as in Experiment 2A, but the five emotional in-

sertions related to the victimization in the crime. Here, the intent was to evoke a sense of empathy in the participants. Gender was again orthogonal to both variables.

Method

Participants. Ninety-six University of Pittsburgh undergraduates (48 men, 48 women), not having participated in any of the previous studies described here, participated in the experiment for partial course credit.

Design and materials. This study followed the same design as described for Experiment 2A. As indicated, however, neither the uncertainty changes nor the emotion manipulations involved the crime itself. The five changes that produced the crime-irrelevant uncertainty were as follows.

1) . . . the earthquake that happened on Tuesday . . . → . . . the earthquake that I believe happened on Tuesday . . . ; 2) . . . windows broken and this led to widespread looting. → . . . windows broken and this most likely led to widespread looting; 3) Because they were concerned about losing their inventory . . . → Probably because they were concerned about losing their inventory . . . ; 4) . . . many shopkeepers stayed in their stores until the police came . . . → . . . many shopkeepers stayed in their stores at least until the police came . . . ; 5) Vince Morelli was one of these shopkeepers, staying in his pharmacy in order to protect his goods. → Vince Morelli was one of these shopkeepers, staying in his pharmacy, we assume to protect his goods.

The emotion condition consisted of these five victim-related changes:

1) Vince Morelli was one of these shopkeepers . . . → Vince Morelli was one of these devoted shopkeepers . . . ; 2) Edward pointed his gun at Vince Morelli → Edward pointed his gun at the shocked Vince Morelli . . . ; 3) He was given the keys. → With trembling hands, he was given the keys; 4) Vince Morelli was not going to let this happen. → Mustering up his courage, Vince Morelli was not going to let this happen; 5) Vince Morelli reached under the cash register . . . → Vince Morelli tried to control his shaking as he reached under the cash register.

As in Experiment 2A, a within-subjects design was achieved by using one condition from each of the four stories so that each participant read all four stories and all four conditions, counterbalanced appropriately. The measures and procedures used here were identical to those used in Experiment 2A.

Results

Guilt ratings. The means presented in Table 5 indicate uncertain crime-irrelevant information had a borderline significance effect on the guilt ratings, $F(1, 352) = 3.72, p < .055$, producing lower guilt ratings than the baseline. This finding suggests that irrelevant uncertainty may have created a general perception for the reader that the entire prosecutor’s statement was uncertain. As shown in Table 5, the baseline condition yielded a higher mean guilt rating than the emotion condition, $F(1, 352) = 4.21, p < .04$. Thus, although the emotional contents emphasized the victim’s plight, lower judgments of guilt were obtained in the emotion conditions. Also, overall, women did provide significantly higher guilt ratings than men, $F(1, 352) = 12.15, p < .001$ (see Table 5).

Story also had a significant effect on guilt judgments, $F(3, 352) = 2.73, p < .04$. The mean of Story 3 differs significantly from the means of Stories 1 and 4. As with Experiment 2A, this effect appears to be driven by Story 3, with its baseline having a relatively salient alternative hypothesis. The only significant interaction of the guilt rating analysis is the Certainty \times Emotion \times Story interaction, $F(3, 352) = 3.67, p < .01$. For the uncertain baseline condition, ratings for the non-emotional baseline are higher than the emotion content condition for Stories 1, 2, and 3, with the reverse holding for Story 4. For the uncertain information condition, ratings are higher with emotional contents in Stories 1 and 2, with the reverse in Stories 3 and 4.

Goodness judgments. The results for the three goodness judgments were reasonably consistent. The baseline condition was rated as having significantly

TABLE 5
Means for the Main Effects of Experiment 2B

	<i>Guilt</i>	<i>Goodness Measures</i>		
		<i>Quality</i>	<i>Convincingness</i>	<i>Evidence</i>
Information certainty				
Baseline	8.74**	3.63***	3.37***	7.00
Uncertain	8.25	3.34	3.07	6.89
Emotion				
Baseline	8.76**	3.58**	3.34**	7.26***
Emotional	8.23	3.39	3.09	6.62
Gender				
Male	8.05****	3.44	3.11*	6.73**
Female	8.94	3.53	3.33	7.15
Story				
1 (Burglary)	8.67**	3.62***	3.45****	7.32****
2 (Hunting)	8.55	3.48	3.17	6.75
3 (Mob attack)	7.90	3.18	2.84	6.18
4 (Wife killing)	8.87	3.67	3.42	7.51

* $p = .06$. ** $p = .05$. *** $p = .01$. **** $p = .001$.

better quality and convincingness than the narrative containing crime-irrelevant uncertain information, $F(1, 352) = 7.37, p < .01$, and $F(1, 352) = 6.14, p < .01$, respectively. However, quality of evidence ratings did not vary significantly ($F < 1$). This finding indicates that individuals separated evidence evaluation from the quality and convincingness judgments and also suggests why the guilt ratings of the baseline and uncertain conditions were of borderline significance. In fact, if evidence was the sole criterion for judgment for the guilt ratings, the difference in these conditions should not have approached significance because evidence was held constant.

The inclusion of victim-related emotional statements produced lower ratings than the baseline for all three goodness ratings, $F(1, 352) = 3.39, p < .07$ for quality; $F(1, 352) = 4.21, p < .04$ for convincingness; and $F(1, 352) = 8.65, p < .01$ for evidence. These findings are in agreement with the guilt rating results, indicating that participants were not convinced by the prosecutor's sympathetic statements about the victim. Instead, it appears that including these statements backfired, as guilt and goodness ratings were lower than they were when these statements were not included.

Gender of participant was not significant for the quality ratings ($F < 1$), but was of borderline significance for convincingness judgments, $F(1, 352) = 3.22, p < .07$, and evidence judgments, $F(1, 352) = 3.72, p < .05$. Women provided higher goodness ratings than did men, consistent with the guilt ratings.

Story was significant for all goodness ratings, $F(1, 352) = 7.37, p < .01$ for quality; $F(1, 352) = 5.27, p < .001$ for convincingness; and $F(1, 352) = 7.66, p < .001$ for evidence ratings. Furthermore, each of these goodness measures produced the same pattern of story differences as the guilt rating data. In all four cases, Story 3 yielded significantly lower ratings than Stories 1 and 4. No other difference was significant. The reasons for this finding are most likely the same as those suggested for the guilt rating findings.

The single significant interaction found in each of the three goodness measures was Uncertainty \times Story \times Gender, $F(3, 352) = 3.04, p < .03$ for quality; $F(3, 352) = 2.81, p < .04$ for convincingness; and $F(3, 352) = 3.68, p < .01$ for evidence. The pattern of means for this interaction is consistent for the convincingness and evidence ratings with one difference in the quality measure. For Stories 2, 3, and 4, the increased uncertainty condition caused ratings to decrease for both men and women, except for men's quality ratings of Story 2. For Story 1, adding uncertainty caused the men to decrease their guilt ratings, but caused women to increase their ratings for all three measures.

GENERAL DISCUSSION

Although Voss et al. (1999) reported that narrative quality, defined in terms of coherence-chronology, influenced jurors' guilt ratings when evidence is held constant, the findings reported here indicate that the role of narrative quality was

related to evidence conditions. These results show that poor evidence produces low guilt ratings, and Klettke and Graesser (2000) showed that contradiction in a narrative also produces low guilt ratings. However, when evidence is either overwhelmingly strong or weak, narrative quality is of relatively little importance. It tends to be of greater importance when evidence is inconclusive, as in the Voss et al. (1999) study. Experiment 1 also specifically showed that a good narrative does not compensate for bad evidence.

If it is assumed that a "juror" in this research constructed a scenario from the prosecutor's statement, the findings suggest that the juror had the ability to construct the scenario even in the disruptive narrative condition. Indeed, the relatively low goodness judgments for the disrupted narrative indicated that, at least for present conditions, the jurors were adept at separating evidential and narrative components. Furthermore, men focused more on the evidence, or at least weighed it more in their guilt judgments, than women. This finding would suggest that scenarios constructed by men may be more skeletal and heuristic and based on evidence, whereas female representations would include more scenario components. This suggestion is consistent with the Meyers-Levy (1989) selective hypothesis and with Kuhn's (in press) distinction of the "satisficing" jurors and the jurors who look at evidence more thoroughly. The view also is in agreement with information processing theories that embrace peripheral and central processing modes (Chaiken, 1980; Petty & Cacioppo, 1986).

There is another possibility regarding the nature of the processing. Schum (1993), following an earlier diagrammatic model by Wigmore (1937), noted that whereas some jurors construct scenarios, others may construct relational models, that is, models not temporal in nature but models that resemble diagrams of ill-structured problem solving (Voss, Greene, Post, & Penner, 1983). It is possible that some jurors, especially men, construct this type of model in their focus on evidence. A take-home lesson from Experiment 1 is that if you have reasonably good or quite good evidence, you should hire an attorney who is a fine narrator; if your evidence is poor, do not go to court.

Experiments 2A and 2B refined and added to the findings of Voss et al. (1999) concerning information uncertainty. Somewhat surprising is the finding that regardless of whether the information uncertainty is related to or irrelevant to the crime, the guilt ratings, compared to the baseline, are about 0.5 units lower and of borderline statistical significance. These findings, as previously noted, suggest that the narrative statements providing uncertainty produce an overall perception of the prosecuting attorney's narrative as indefinite. Intuitively, the crime-related statements should have produced a greater effect than the irrelevant statements because the crime-related statements should raise doubts about the defendant's guilt, whereas the irrelevant uncertainty should not.

The goodness ratings indicated that although having uncertainty in the crime-related narrative led to the narrative being judged as poor, the influence on the guilt ratings, which was not as strong, suggests the jurors regarded the narrative

characteristics and evidence as separate, although related, narrative components. This notion is especially shown in the crime-irrelevant uncertainty condition, in which the goodness results were essentially the same as in the crime-related narrative, except that the evidence quality ratings were not different from the baseline. This finding indicates that the jurors in the uncertain crime-irrelevant condition quite readily separated the evidential and uncertain crime-irrelevant components. Again, the data do not tell us whether the jurors constructed a relational structure. A take-home lesson is that it is desirable to hire an attorney who is a clear and concise speaker, not given to qualifiers or hedges that may suggest doubt about the facts surrounding the crime.

The inclusion of crime- or defendant-related emotional statements in the prosecuting attorney's narrative yielded higher goodness ratings for the narrative than those found in the baseline condition, but there was no main effect of emotion on guilt ratings, despite the fact that evidence was rated as better than the baseline of the emotion condition. However, men did produce lower guilt ratings, compared to the baseline for Stories 1, 2, and 4, whereas women produced lower guilt ratings only on Story 3. Conversely, women produced higher guilt ratings on Stories 1, 2, and 4, whereas men produced higher ratings only on Story 3. A possible explanation for this interaction is that in most cases, as suggested by Meyers-Levy (1989), men relied more heavily on the evidence cue in making their judgments. They may have found the emotional statements distracting and reacted against them, thereby resulting in lower guilt ratings than those found in the baseline. Women, who are presumed to give more comprehensive consideration to the information in the stories, apparently took the emotional statements into account, producing higher ratings than the baseline in Stories 1, 2, and 4. As to the Story 3 results, whereas the negative emotional statements in Stories 1, 2, and 4 referred to personal characteristics of the defendant, the emotional statements in Story 3 were more pertinent to the situation surrounding the crime itself, including the evidence. Thus, men apparently were influenced by the emotional statements related to evidence; women seem to be more sensitive to criminal-related emotional statements. Thus, the data do indicate that the emotional information is incorporated into the narrative representation, but the extent to which it influences guilt ratings depends on the story, the nature of the emotion described in the story, and the gender of the juror.

When the narrative contains victim-related emotional expressions, guilt ratings are below the baseline, as are goodness ratings. Such emotion worked to oppose the idea that victim-related emotional statements would evoke empathy and produce higher guilt ratings than the baseline.

The incorporation of gender into these experiments produced the following tentative conclusions. Men responded more extremely to good or poor evidence than women. As previously mentioned, this finding supports hypotheses suggesting men do more heuristically focused processing than women, who do more comprehensive processing. Men and women also differentially respond to the

specific emotional contents of particular stories. In general, men are influenced more if the emotional content can be related to evidence, whereas women are more influenced if the emotional content refers to the defendant. Although in Experiment 2B women, overall, provided higher guilt ratings than men, the failure to obtain particular significant interactions indicated there was no evidence that women were influenced by victim-related emotional statements.

Story effects occurred in all experiments despite attempts to equate the stories along characteristics related to the research. The fact that story effects were nevertheless different and that they interacted with the variables under study meant that we could only apply an a posteriori explanation regarding these effects. The results do suggest that guilt ratings may be a function of the extent to which an alternative hypothesis for the crime may be plausible. Also suggested is that gender interacts with the type of story and its emotional content. To study such inferences more thoroughly, simultaneous manipulation of the individual and story contents would be necessary.

In sum, narratives may be used in a courtroom to provide support for a claim about a defendant's guilt or innocence. Jurors do respond to the evidence as stated in the narrative, but factors such as uncertain statements and emotional expressions can influence the decision making of the jurors in these studies. We have seen here that individual differences of the jurors also play a role (Kuhn, in press), among them gender. This factor may produce differences in types of processing, with more specific heuristic processing or more comprehensive processing representing extremes of a range of many differences.

ACKNOWLEDGMENT

We thank Deanna Kuhn, Jennifer Wiley, and Ronan Bernas for their helpful comments.

REFERENCES

- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39, 752–766.
- Foucault, M. (1972). *The archeology of knowledge* (A. Sheridan Smith, Trans.). New York: Pantheon. (Original work published 1969)
- Gernsbacher, M. A., Hallada, B. M., & Robertson, R. W. (1998). How automatically do readers infer fictional characters' emotional states? *Scientific Studies of Reading*, 2, 271–300.
- Klettke, B., & Graesser, A. C. (2000, July). *Coherence and evidence in testimony evaluation in incest narratives*. Paper presented at the 10th annual meeting of the Society for Text and Discourse, Lyon, France.
- Kuhn, D. (2001). How do people know? *Psychological Science*, 12(1), 1–8.

- Leinhardt, G., Stainton, C., Virji, S. M., & Odoroff, E. (1994). Learning to reason in history: Mindlessness to mindfulness. In M. Carretero & J. F. Voss (Eds.), *Cognitive and instructional processes in history and the social sciences* (pp. 131–158). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9, 111–151.
- Meyers-Levy, J. (1989). Gender differences in information processing: A selectivity interpretation. In P. Cafferata & A. M. Tybout (Eds.), *Cognitive and affective responses to advertising* (pp. 219–260). Lexington, MA: Heath.
- Pennington, N., & Hastie, R. (1993). The story model for juror decision making. In R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making* (pp. 192–221). Cambridge, England: Cambridge University Press.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–205). New York: Academic.
- Schum, D. A. (1993). Argument structuring and evidence evaluation. In R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making* (pp. 175–191). Cambridge, England: Cambridge University Press.
- Trabasso, T., van den Broek, P., & Suh, S. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes*, 12, 1–25.
- van Eemeren, F. H., Grootendorst, R., & Snoecele Henkemans, F. (1996). *Fundamentals of argumentation theory*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem solving skill in the social sciences. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research theory* (Vol. 17, pp. 165–213). New York: Academic Press.
- Voss, J. F., Wiley, J., & Sandak, R. (1999). On the use of narrative as argument. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative, comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 235–252). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Wigmore, J. H. (1937). *The science of judicial proof as given by logic, psychology, and general experience, and illustrated in judicial trials* (3rd ed.). Boston: Little, Brown.

APPENDIX A BASELINE, STORY 2

Prosecutor's Statement

Paul Ryan and Rob Brown drove to DuBois together in Rob's truck. For years they had driven there on the first Saturday of deer season. They had packed the truck on Friday night so they could get an early start. Rob had looked forward to this hunt for the entire year. This time, however, Rob Brown was shot and killed by Paul Ryan. According to Paul, Rob was walking through the woods about 40 yards away when Paul saw the head and antlers of a deer to his right. He raised his gun, aiming at the deer, and fired. He claims that, being a poor shot, his shot missed the deer, but that he saw something else fall. He said he ran to see what he had hit, and found Rob lying in the grass, dead.

Paul says that he immediately went to a nearby cabin where he contacted the police. He claims that this was all a terrible accident. However, the police investigation concluded that, while there were deer tracks in the area where Paul and

Rob had been walking, the area had only a few trees and a little brush and thicket, and that Rob was quite likely fully visible to Paul. Furthermore, during the investigation, Paul's wife, Rachel, testified that she and Rob had been having an affair for over a year and a half. Mary Jacoby, a friend of Rachel Ryan's, testified that she, Mary, had thought there was an affair and had told Paul about her suspicions just prior to the hunting trip. Since Paul had just discovered that Rob had been having an affair with his wife, he had a motive. Furthermore, since the police have established that Paul was quite likely able to see Rob when he was shot, the death of Rob Brown was a premeditated murder and not a freak accident. Paul should be found guilty of first-degree murder.

Experiment 1: Evidence Manipulations

1) . . . (the area had only a few trees and a little brush . . . Rob was quite likely fully visible to Paul → . . . the area had a number of trees and quite a bit of brush . . . making it difficult for Paul to see Rob; 2) Rachel testified that she and Rob had been having an affair for over a year → Rachel testified that before she and Paul were married, she and Rob went together for over a year; 3) Mary had thought there was an affair and had told Paul about her suspicions → Mary had known about Rachel and Rob going together before Rachel's marriage to Paul; 4) Since Paul had just discovered that Rob had been having an affair with his wife, he had a motive. → Since Rob had gone with Rachel some years ago, Paul had a motive to shoot Rob; 5) the police have established that Paul was quite likely able to see Rob → the police could not establish that Paul was able to see Rob.

Experiment 2A: Uncertain Information Manipulations

1) . . . and that Rob was quite likely fully visible to Paul. → . . . and that Rob could have been fully visible to Paul; 2) . . . Rob had been having an affair for over a year and a half. → . . . Rob had been having sort of an affair for over a year and a half; 3) Since Paul had just discovered that Rob had been having an affair . . . → Since Paul had just discovered that Rob probably had been having an affair . . . ; 4) . . . he had a motive. → he also may have had a motive; 5) . . . the police have established that Paul was quite likely able to see Rob when he was shot . . . → . . . the police have established that Paul might have seen Rob when he was shot. . . .

Experiment 2A: Emotion Manipulations

1) This time, however, Rob Brown was shot and killed . . . → This time, however, Rob Brown was brutally and mercilessly shot and killed . . . ; 2) . . . but that he saw something else fall. → . . . but that he saw something else fall violently to the ground; 3) . . . Rob lying in the grass, dead. → . . . Rob lying in the grass, dead, his blood splattered all over the woods; 4) . . . she and Rob had been having an affair

for over a year . . . → . . . she and Rob had been having a steamy, illicit love affair for over a year . . . ; 5) . . . the police have established that Paul was quite likely able to see Rob when he was shot . . . → . . . the police have established that Paul was able to see Rob, his innocent and unsuspecting target, when he was shot . . .

Experiment 2B: Uncertain Information Manipulations

1) Paul Ryan and Rob Brown drove to . . . → Paul Ryan and Rob Brown usually drove to . . . ; 2) . . . drove to Dubois together . . . → drove to Dubois or some other place together . . . ; 3) . . . together in Rob's truck. → together in Rob's truck or possibly Paul's car; 4) . . . there on the first Saturday of deer season. → . . . there on the first Saturday or at least sometime during deer season; 5) They had packed the truck on Friday . . . → They had probably packed the truck on Friday. . . .

Experiment 2B: Emotion Manipulations

1) . . . there on the first Saturday of deer season. → . . . there on the first Saturday of deer season, always having a good time laughing at Rob's jokes; 2) They had packed the truck . . . → Excited about the hunt, Rob suggested they pack the truck . . . ; 3) Rob had looked forward to this hunt . . . → Rob had enthusiastically looked forward to this hunt . . . ; 4) . . . to this hunt for the entire year. → . . . to this hunt with his best friend for the entire year; 5) According to Paul, Rob was walking through the woods . . . → According to Paul, alert and cautious, Rob was walking through the woods. . . .

APPENDIX B BASELINE, STORY 3

Prosecutor's Statement

The victim, Roger Wilson, was hit by a baseball bat and died. This is how the crime took place. Roger Wilson had dropped off his co-worker, Susan Walker, at her home. He then drove down Crawford Road in order to get onto the freeway. As he was driving, he saw a small girl run out from behind a parked car. Before he could stop, his right fender hit her and she fell to the ground. Roger quickly got out of his car and checked to see if the girl was seriously hurt. She was not. Roger turned to call the police, but a group of teenagers began to push him around. Then one of the teenagers took a baseball bat and hit Roger with it. Roger fell over and died.

A few minutes later police arrived on the scene and interviewed those present. A person living across the street had witnessed the events, but she could not iden-

tify who struck Roger. Then the police discovered a blood-covered bat in the rear seat of a car parked nearby. The car belonged to Matthew Moran, the older brother of the girl that Roger had hit. Later analyses revealed that the blood was that of the victim and that the victim's hair was also on the bat. The bat also had fingerprints on it, but the fingerprints were smudged and could not be identified as those of Matthew Moran. Matthew claims someone else threw the bat into his car through the open window, but the police didn't mention an open window in their report. We have also learned that Matthew often became angry when he felt someone was mistreating his little sister. This served as the motive to hit the victim.

The evidence therefore indicates that Matthew Moran used the bat to kill Roger Wilson. He had the motive, the means, and the opportunity.

Experiment 1: Evidence Manipulation

1) the blood was that of the victim . . . → the blood could not be identified as that of the victim; 2) the victim's hair was also on the bat. → it was unclear whose hair was on the bat; 3) the police didn't mention an open window in their report. → police did mention in their report that the car windows were open; 4) Matthew often became angry . . . → Matthew typically was a caring person and not temperamental; 5) This [his anger] served as the motive to hit the victim. → seeing his sister get hit by the car probably served as the motive to hit the victim.

Experiment 2A: Uncertain Information Manipulation

1) Later analyses revealed that the blood was that of the victim . . . → Later analyses revealed that the blood was probably that of the victim . . . ; 2) . . . and that the victim's hair was also on the bat. → . . . and that what looked like the victim's hair was also on the bat; 3) . . . fingerprints were smudged and could not be identified as those of . . . → . . . fingerprints were smudged and could not be identified completely as those of . . . ; 4) We have also learned that Matthew often became angry . . . → We have also learned that Matthew may have often become angry . . . ; 5) This served as the motive to hit the victim. → . . . and there is a chance this served as the motive to hit the victim.

Experiment 2A: Emotion Manipulation

1) . . . but a group of teenagers began to push him around. → . . . but a mob of rowdy, yelling teenagers began to push him around; 2) . . . took a baseball bat and hit Roger with it. → . . . took a baseball bat and viciously struck Roger with it; 3) Roger fell over and died. → Roger fell over, his face getting covered with blood as he collapsed, and died; 4) A person living across the street had witnessed the events . . . → A person living across the street had witnessed the brutal and disturbing events . . . ; 5) . . . Matthew often became angry when he felt someone was

. . . → . . . Matthew often became angry, losing his temper and becoming violent when he felt someone was. . . .

Experiment 2B: Uncertain Information Manipulations

1) Roger Wilson had dropped off his co-worker . . . → Roger Wilson had apparently dropped off his co-worker . . . ; 2) He then drove down Crawford Road in order to get onto the freeway. → He then drove down Crawford Road, probably to get onto the highway; 3) As he was driving, he saw a small girl . . . → As he was driving, he thought he saw a small girl . . . ; 4) A few minutes later police arrived . . . → Witnesses thought it was a few minutes later when police arrived . . . ; 5) A person living across the street . . . → A person who probably lives across the street. . . .

Experiment 2B: Emotion Manipulations

1) He then drove down Crawford Road in order to get onto the freeway. → He then drove carefully down Crawford Road in order to get onto the freeway; 2) . . . he saw a small girl run out from behind a parked car. → . . . he saw a small girl run out from behind a parked car and it frightened him; 3) Roger quickly got out of his car . . . → Shaking and perspiring, Roger quickly got out of his car . . . ; 4) . . . checked to see if the girl was seriously hurt. → . . . gently but carefully checked to see if the girl was seriously hurt; 5) Roger turned to call the police . . . → Relieved, Roger turned to call the police. . . .

APPENDIX C BASELINE, STORY 4

Prosecutor's Statement

It all began when Raymond Hammond saw his ex-wife, Myrna Ramsey, walking along the jogging trails near Bridge Street. She had just moved from Castle City. Five years earlier, Raymond had swindled a large amount of money from her when she worked at an insurance company there. Now Raymond was afraid that she would tell his current wife about these shady dealings. When he saw Myrna stop to talk to two people who were walking their dogs, Raymond approached her. Raymond told Myrna that he would kill her if she talked with his current wife Connie. He then left in a rush, crossing through the busy traffic on Forward Street.

Three days later Myrna did call Connie and arranged to meet her. Raymond was home when Myrna called and he overheard the conversation from the bedroom. He had heard where Myrna and Connie were to meet, at the Midway Restaurant, and he drove to that destination a half-hour before the appointed time. He

waited in a sheltered doorway across the street from the restaurant until he saw Myrna approaching. When he saw her, he grabbed her purse, took out his gun, and shot her. He then quickly got in his car and drove away. Later, two eyewitnesses were sure that his car was the one they saw driving away from the scene. Four hours later, the police picked him up at his home. During the intervening 4 hours, Raymond Hammond attempted to get rid of the murder weapon by driving his car to the river and throwing the gun in the water. Later, the gun was recovered and identified as his. Having been at the river bottom, the gun had no fingerprints on it. Raymond claims that he had lost that gun several months before and didn't know he should report it missing. In light of the evidence presented, Raymond Hammond should be found guilty.

Experiment 1: Evidence Manipulation

1) Raymond was home when Myrna called . . . → Raymond was not home when Myrna called; 2) . . . he drove to that destination a half-hour before the appointed time. → . . . a man resembling Raymond Hammond was seen in the area a half-hour before the appointed time; 3) two eyewitnesses were sure that his car was the one they saw driving away → two eyewitnesses indicated they saw the car but could not identify the driver; 4) During the intervening 4 hours, Raymond Hammond attempted to get rid of the murder weapon by driving his car to the river . . . → During these 4 hours, Raymond Hammond said he was having lunch and watching a movie; 5) the gun was recovered and identified as his. → a gun was recovered but its owner is not identified.

Experiment 2B: Uncertain Information Manipulation

1) . . . walking along the jogging trails near Bridge Street. → . . . walking along the jogging trails probably near Bridge Street; 2) She had just moved from Castle City. → She may have just moved from Castle City; 3) When he saw Myrna stop to talk to two people . . . → When he saw Myrna stop, possibly to talk to two people . . . ; 4) He then left in a rush, crossing through . . . → He then left in a rush, probably crossing through . . . ; 5) Three days later Myrna did call . . . → Around 3 days later Myrna did call. . . .

Experiment 2B: Emotion Manipulation

1) . . . walking along the jogging trails near Bridge Street. → . . . walking peacefully along the jogging trails near Bridge Street; 2) She had just moved from Castle City. → Feeling like she was starting a new life, she had just moved from Castle City; 3) . . . when she worked at an insurance company there. → When she worked at an insurance company there and she felt that she had been manipulated; 4) When he saw Myrna stop to talk to two people . . . → When he saw Myrna stop

to cheerfully talk with two people . . . ; 5) Raymond told Myrna that he would kill her . . . → Raymond told Myrna, shocked to see him, that he would kill her. . . .

Experiment 2A: Uncertain Information Manipulation

1) . . . when Myrna called and he overheard the conversation . . . → . . . when Myrna called and he might have overheard the conversation . . . ; 2) He waited in a sheltered doorway . . . → Apparently, he waited in a sheltered doorway . . . ; 3) When he saw her . . . he took out his gun, and shot her. → When he saw her, he presumably took out his gun, and shot her; 4) Later, two eyewitnesses were sure that his car . . . → Later, two eyewitnesses were fairly sure that his car . . . ; 5) Raymond Hammond attempted to get rid of the murder weapon . . . → . . . Raymond Hammond evidently attempted to get rid of the murder weapon. . . .

Experiment 2A: Emotion Manipulation

1) . . . at the Midway restaurant, and he drove . . . → at the Midway restaurant, and with murder in his heart he drove . . . ; 2) When he saw her . . . he took out his gun . . . → Fuming with anger when he saw her, he took out his gun . . . ; 3) He then quickly got in his car and drove away. → As she sunk to the ground in a pool of blood, he then quickly returned to his car and drove away; 4) . . . the one they saw driving away from the scene. → . . . the one they saw driving in a reckless fury away from the scene. 5) . . . the police picked him up at his home. → . . . the police picked him up at his home after he threatened to attack the officers.