A study examined how people's level of moral judgment affects their memory for narratives containing moral information. Subjects, 69 eighth-grade students from a private school, 53 eighth-grade students from a public school, and 49 undergraduates enrolled in introductory psychology courses at the University of Minnesota, read four narratives containing embedded moral arguments. The Defining Issues Test (DIT) scores were used as an indicator of moral judgment development and were correlated with moral judgment recall. Results indicated that: (1) higher DIT scores (indicative of structures of moral reasoning) were significantly related to higher principled argument recall (indicative of moral comprehension); and (2) higher DIT scores were significantly related to better overall memory (general reading comprehension). Findings suggest that there is an influence of structures on recall. (Four graphs of data are included.) Contains 30 references. (RS)
MORAL JUDGEMENT AND TEXT PROCESSING
Darcia Narváez, University of Minnesota
206 Burton, 178 Pillsbury, Mpls MN 55455
Discourse processing has focused primarily on the processing of written texts.

Research in text comprehension has ascertained that narrative events are recalled best when they conform with two conditions. One, the event falls on a causal chain which runs from the beginning to the end of the text. Two, the event has many causal connections with other events in the text. Both conditions lead to better recall than for events that have few connections or that are not on the causal chain. Events with causal chain status and with a greater number of connecting relations to other events in the text are more critical to the understanding of other events in the story because they form the basis for the reader's construction of a coherent representation of the text (Trabasso & van den Broek, 1985; Trabasso & Sperry, 1985; van den Broek, 1990; Graesser & Clark, 1985). When tasks related to the text are undertaken, it is assumed that the reader consults this representation.

Bower & Cirilo (1985) have suggested two kinds of text processing that influence the reader's representation. First, a data-driven processing activates a high-level structure or schema. Second, after the high-level structure is activated, it in turn activates accompanying subschemata or conceptually-driven processing, i.e., processing guided by expectations. Subsequent processing is an interaction between the two types. When texts are inconsistent with the high-level knowledge structures invoked, readers will poorly understand and misrecall and even distort memory to fit with their schematic form as in Bartlett's seminal work (1932) with 'The War of the Ghosts' story.

Differences in comprehension as a result of varying expectations are evident in several situations. Perspective plays a role in what is represented and later remembered.
Pichert and Anderson (1978) demonstrated that reading a description of a home either as a potential burglar or a potential home buyer influences the content of memory one has for the characteristics of the home. Anderson, Reynolds, Schallert & Goetz (1977), using a paragraph that could be interpreted as a wrestling match or as an attempted prison escape, postulated that a subject's schemata provides the interpretive framework for the understanding of a discourse. Physical Education majors interpreted the story as wrestling while non-majors interpreted it as a prison escape. Bartlett (1932) was the first in this century to provide evidence for the influence of cultural scripts on recall. Harris, Lee, Hensley & Schoen (1988) found that routines from another culture were increasingly misrecalled over time by those from a different culture, indicating schematic influence on retrieval processes. Thus, individual differences in comprehension may result from conceptually-driven processing which affects the coherence of a text.

Coherence is an important feature of comprehension. Two kinds of coherence have been postulated: an objective, or within-text, coherence and a subjective, or within-person, coherence. Within-text coherence plays an important role in a wide range of comprehension situations varying as far as jury decision-making (Pennington & Hastie, 1986). The within-text coherence can be objectively measured with, for instance, co-referential cohesion (Kintsch & van Dijk, 1978) or counterfactual reasoning for causal cohesion (Trabasso & Sperry, 1985) and is related to recall and importance ratings of textual events. This can be measured for simple written texts, tv shows and films, by forming the corresponding networks of events presented. However, these measures are less easily applied to more complex texts and films, where multiple perspectives are possible, such as moral events, and,
especially, to real-life events since real-life events require some kind of structural imposition for interpretative understanding. Unstructured stimuli are often not perceived let alone comprehended (Neisser, 1967).

Thus, the second kind of coherence, subjective coherence, is necessary to mediate the objective coherence. When data-driven processing does not activate a schema for organization within the reader because, for instance, the text's statements are not constraining enough, comprehension may not ensue (Bransford & Johnson, 1972). People often attempt to impose structure when there clearly is none as when reading unrelated sentences or when observing random events (Kelley, 1973; White, 1988). The use of particular causal or referential relations seems to vary by culture and task as cited earlier. Subjective coherence may vary, then, according to cultural schemata, task, expertise, as well as the specific text.

Morality has traditionally played a large role in the generation and understanding of stories, movies, and plays. Little, if any, research has addressed the critical features that define coherence in a moral narrative. This research is an attempt to do so.

Research in morality has focused mostly on moral reasoning and has been performed primarily by cognitive developmentalists (Kohlberg, 1969; Rest, 1979) who base their view of cognitive development on the work of Jean Piaget. Within this view, cognitive skills build on earlier skills; specifically, certain cognitive changes are prerequisite to changes in moral reasoning (Rest, 1983). Like Piaget's cognitive stages, moral judgment stages are sequential and describe moral judgment development in terms of successive transformations of the concepts of justice. Consistent and reliable effects have been found for increasing use of complex stages due to age and, primarily, amount of higher education or similar experience.
It has been demonstrated that one's prior mental constructs and schemas can bias perception (Bransford & McCarrell, 1974; Neisser, 1967; Nelson, 1986). The initial representation of an event, as modified by this perceptual bias, constrains comprehension and subsequent recall (Bartlett, 1932; Pichert & Anderson, 1978). It is postulated here that schemas in moral reasoning are not scripts with slots to be filled but rather organizational structures that order perception and, thereby, subsequent related output. This work attempts to measure whether one's emerging moral cognitive structures, as measured by moral reasoning preference, influence one's comprehension, as measured by recall.

This research approaches one aspect of moral discourse processing, the influence of moral judgment schemas on text processing. It examines the effects of theory-driven, schematically-based processing on memory for events in a text. Specifically, this research examines how a person's level of moral judgment affects his/her memory for narratives containing moral information. Of particular interest is the interaction of judgment, age, and recall of moral arguments embedded in textual information.

Two kinds of comprehension will be studied, moral reasoning comprehension and general reading comprehension. Unlike in other research with texts where coherence is measured for different stories with similar subjects (e.g., Trabasso & Sperry, 1985), this research looks at how coherence varies for the same stories in relation to individual cognitive differences. The individual cognitive difference of interest is moral reasoning development. There is abundant evidence that moral reasoning is a cognitive activity that builds upon but is separate from intelligence and specific mental skills like calculus (Rest, 1979). Unlike in most other research involving moral reasoning where spontaneous verbalizations of reasoning
in response to dilemmas are measured, memory for moral arguments (moral judgment stage arguments) embedded within narratives is measured in these investigations.

One study that used a recall measure is Rest, Turiel & Kohlberg (1969). They measured recall for written moral 'advice' given by several people arguing different positions on a moral dilemma. There was highest recall for arguments one stage below the subject's own stage and decreasing correct recall for arguments above the subject's stage. They argue that this is an effect of comprehension not preference. They found a general tendency of the subject to distort the moral arguments to his or her own level or one stage below, along with a strong likelihood of distorting disliked advice to one stage below.

In this study, there are some similarities to the Rest et al. design along with some critically important differences. As did Rest et al., a recall measure was used and subjects' responses to moral judgments were recorded. Unlike Rest et al., the moral stage arguments are presented implicitly in narratives rather than explicitly in separate statements. In other words, the events of interest are presented in a stream of information, as happens in real life; the subject's attention is not directed to the arguments as in Rest et al. but allowed to focus freely. Additionally, the stage arguments embedded in each story vary unsystematically, i.e., every stage is not represented in every story since not every stage would fit naturally into a story; this is an ecologically valid feature (Rest, 1979). Whereas Rest et al. (1969) measured both preference and comprehension, here only explicit comprehension is measured but more thoroughly; recall for critical and uncritical events, and for moral arguments at various stages were scored for each story and for each subject.

Perception, as biased by existing schemas, frames one's interpretation of and memory
for events, be they real-life events or events within a narrative. In this study, the events include moral arguments within a story. In the comprehension of narratives involving moral dilemmas, it is suggested that memory for the narratives will be determined by an interpretive structure based on the reader’s level of moral judgment. The subjects’ level of moral judgment will be measured by Defining Issues Test (DIT), an objective measure of moral judgment (Rest, 1979). It requires subjects to rate the importance of considerations, representing moral judgment stages, in making a moral decision.

Subjects will read four narratives containing embedded moral arguments. It is postulated that subjects will recall understandable arguments but distort or assimilate downwards the arguments that are beyond their comprehension. The Defining Issues Test "P" (Principled thinking) score will be used as an independent variable, as an indicator of moral judgment development, to be correlated with moral argument recall. It is postulated here that the students who prefer higher stage arguments on the DIT will recall those same-level arguments when they recall the narratives. For low moral comprehenders moral arguments that are beyond comprehension will either not be recalled, dismissed as incomprehensible, or will be reinterpreted at a simpler level of understanding.

**EXPERIMENTAL HYPOTHESES**

**HYPOTHESIS 1: Age differences in P score**

Older students will obtain higher DIT P scores. University students will prefer more principled reasoning on the DIT than the 8th grade students.

**HYPOTHESIS 2: Critical event recall**

Subjects will recall a significantly greater number of critical events, those with three or more
causal connections in the mental representation, than non-critical events, those with less than three causal connections.

**HYPOTHESIS 3: Age differences in critical event recall**

Older subjects (college age) will recall more of the critical story events than younger (junior high) subjects.

**HYPOTHESIS 4: Age differences in overall recall**

Older subjects have better overall memory (content memory).

**HYPOTHESIS 5: Recall for principled arguments**

Subjects with higher P scores will recall more of the higher stage moral arguments in the stories.

**METHOD**

**Subjects.** The subjects were 161 8th grade students from a private (N=80) and a public school (N=81) in Minnesota suburbs and 62 undergraduates enrolled in introductory psychology courses at the University of Minnesota. Due to uncompleted tasks or failing the DIT's consistency check, the final totals were 69, 53 and 49, respectively.

**Materials.** There were two sets of materials. The first involved four stories written by the researcher. (See Appendix A for the stories.) Each story concerned a situation in which the protagonist had to make a moral decision and was considering many options (representing different moral stage levels). Within each story were embedded several different stage-level arguments. Good narrative flow did not allow every stage in every story. "Tom, the manager" concerns a manager who had hired a troubled nephew at his sister's (the mother's) request. As the nephew's performance deteriorates over time, despite
his uncle's supportive aid, Tom debates whether or not to fire him. "Penelope and the check" is about an impoverished woman who receives an insurance check in the mail which is for $2000 instead of $200, with the mistake only on the check, not on the stub. She wrestles with cashing or returning the excess money. "Watching the game" is about Bob and friends who reach the ticket booth as the tickets for the big game run out. On the way back to the cars, one of them sees an open, unsupervised stadium door. They argue about the option to go in for free. "Sara and the demonstration" is about a young woman who arrives home from an exhausting week of work who is invited to an illegal demonstration against a weapons plant by her lifelong best friend. They discuss the merits of demonstrating in this situation. Sara is left to decide whether to possibly save their friendship by going or not.

After reading all the stories, a subject received a set of questions to stimulate recall for each story. The subject answered questions on each story in the same order that they were read. Each story’s questions were the same except that names were changed. For example, "Tom, the manager" used the following questions. 1) "Describe the major events of the story." This question was asked to stimulate reinstatement and reconstruction of the story according to the subject’s own schemata. 2) "What were Tom’s considerations in making a decision?" This question was asked to find out which moral arguments were recalled and whether additional moral arguments were included in the individual’s schemata. The last two questions were asked to find out what the individual would choose and which rationale of those listed he or she would select. 3) "What do you think Tom finally decided?" 4) "Why do you think he would decide that?" Answers to questions one and two were scored for general content recall and for moral argument recall.
The second set of materials was the Defining Issues Test (DIT) which is an objective, paper and pencil measure of moral judgment preference. The test consists of six moral dilemma vignettes. After reading each vignette, the subject rates the importance of and ranks a list of concerns one might have in that particular situation. The Principled or "P" score is the most valid and widely used index (Rest, 1993). It is a weighted sum of the principled reasoning, i.e., postconventional Stages 5 and 6 in the Kohlberg scheme (which emphasize the importance of principles in making moral judgments) preferred by the subject. Test-retest reliability ranges between .70 and .80 for the P score. Internal consistency as measured by Cronbach's Alpha has the same range in various studies (Rest, 1990). The DIT typically takes from 35-50 minutes to complete.

Scoring. The stories were each parsed into propositions that constitute events in the broad sense, much like Warren, Nicholas & Trabasso, (1975). A scoring system was devised for the four stories whereby moral stage arguments, general content memory, non-textual elaborations and errors were scored using a gist criterion, where either a paraphrase or exact wording qualified as a correct answer. Scores were obtained for moral argument recall, critical event recall and noncritical event recall. Critical events were those that were connected to, by being causally necessary in the circumstances for (Trabasso & van den Broek, 1985), three or more other events in the story. Objective coherence, the within-text coherence was diagrammed by drawing each event node in its relation to other events; critical events formed the objective causal chain from the beginning to the end of each story.

Twenty percent of the protocols were co-scored by another judge. Reliability was computed on these protocols as kappa=95%. Disagreements were discussed and resolved.
The entire group of protocols was then rescored by the researcher. The DITs were scored by optical scanner and computer software at the Center for the Study of Ethical Development.

Procedure. Subjects had two basic tasks. First, they were given the four stories to read and recall. The subjects read the stories one after the other. The instructions were to 'read for understanding.' The order of the stories was counterbalanced among subjects. When they finished reading, subjects exchanged the stories for a question package which included questions for each story in the same order in which the subject had read the stories. Subjects were given unlimited time to complete the task. Most finished in less than one hour.

After the story task, subjects took the Defining Issues Test. Because of age differences in attention span, the 8th grade students took the DIT one to two weeks later while the college students took the DIT immediately after the story task. Most students completed the DIT within 45 minutes.

RESULTS

(See tables)

Hypothesis 1: Age differences in the DIT P score were evident. University students obtained higher P scores. The college average was 38.3 and the 8th grade average was 27.2. The t-test was significant at p < .0001. The mean for the college students was slightly below average for their age group (M=42) while the mean for the 8th grade students was above average (M=21.7).

Hypothesis 2: Results similar to those with non-moral laboratory stories were
obtained regarding critical event memory. Subjects recalled a significantly greater number of general critical events, those with three or more causal connections in the objective representation, than non-critical events, those with less than three causal connections. These results were significant for the whole group and for the age groups alone. The whole-group critical memory average was 32% while the noncritical average was 11%. The t-test was significant at p < .0001. The 8th grade mean for critical event recall was 29% and 9% for noncritical events. The t-test was significant at p < .0001. For the college students, critical event recall averaged 40% and noncritical, 14%. The t-test was significant at p < .0001.

Hypotheses 3: There were age differences in critical event recall. University students recalled, on average, more of the critical story events, 40%, than the 8th graders at 29%. The t-test was significant at p < .0001.

Hypothesis 4: Older students had better overall content memory. The college average was 21% while the 8th grade average was 14%. The t-test was significant at p < .0001.

Hypothesis 5: Those who had higher P scores had better recall for higher stage moral arguments in the stories. Each age group was divided into terciles according to DIT Principled score. The low group included those who scored at the junior high level or lower, the middle group contained those who scored between junior high and college level while the high group was composed of those who scored above the college level norm. A two-way ANOVA was run. A main effect for age (p < .0001) and a main effect for P-score group (p < .0002) were obtained. The interaction was also significant.
DISCUSSION

Two kinds of comprehension were measured, the relatively stable structures of moral reasoning and the recently created structures of story recall which involve, most importantly, critical event recall.

Results indicate that there is an influence of structures on recall. Higher DIT scores (indicative of structures) were significantly related to higher principled argument recall (indicative of moral comprehension). This is not to say that they are equivalent. In a separate analysis (Narváez, 1993), 8th grade scores were correlated with scholastic rank on two achievement subtests, vocabulary and reading comprehension. The analysis provided evidence for the separate arenas of comprehension, reading and moral. Cognitive ability was a prerequisite but not sufficient for high scores in moral reasoning.

Additionally, higher DIT scores (structures) were significantly related to better overall memory (general reading comprehension). It is postulated that a greater number of particular causal links were activated for the high comprehenders than for low scorers, who had fewer nodes activated overall and whose links were less differentiated. For the high scorers, it is argued that the respective material presented was, on the whole, more coherent to them than to low scorers. It is postulated that since the embedded moral arguments intertwined with the general story information, these were together, on the whole, more strongly causally interrelated and clearly represented in the memory trace of the high scorers. Further research is required to validate this conclusion.
REFERENCES


APPENDIX A

TOM, THE MANAGER

Tom was having another sleepless night. He quietly crawled out of bed for the fourth time, grabbed his robe and walked to the bathroom. As he stared at the exhausted face in the mirror, he thought about his dilemma for the hundredth time that night.

He was the manager of a store which was part of a department store chain with branches all over the country. It was challenging work which he greatly enjoyed. He had given a job to his nephew, Freddie, because Freddie’s mother, Tom’s sister, had pleaded with him to do so. Freddie had been in a lot of trouble at school and needed a fresh start. Tom liked his older sister and wanted to help her out.

After about three weeks, Tom noticed that Freddie was not doing a good job. Freddie seemed unwilling to do anything. He wouldn’t stock the shelves, he was reluctant to help customers – he would even arrive late.

Several times, Tom had tried to straighten Freddie out. He gave him special instructions and encouragement about how to do his job. He gave him tips on helping customers, techniques for stocking shelves, shortcuts on doing inventory. But nothing seemed to help. Freddie didn’t change. The only thing that Freddie had ever done right was when he had been sent on a special errand to pick up a regional director from the airport. But that was back in the first week Freddie was employed.

Just that morning Freddie had arrived two hours late. They had a big sale going so it was a hardship on his coworkers. They had worked hard trying to serve all the customers but it had been too much. Tom heard a few of them complaining as they punched out and headed for home. Tom pulled out the seventh grey hair that night. He shut off the light and went to the den.

He sank into the easy chair. If only he could get some sleep. He had been tossing and turning for the past two weeks and feeling drained most of the time. He kept thinking.

Freddie just wasn’t doing his job. And he wasn’t improving. As Tom pondered the difficulty, he realized that his position in the company and his special responsibilities were designed to create the greatest benefit for the most people – for the customers, the workers and the investors. It was a necessary part of his job to rid the company of unproductive employees. It was a policy that he fully supported. In fact, Tom would not want to work for a company that didn’t have work standards. Furthermore, Tom was working on the assumption that he and everyone else in the company tacitly agreed with this policy. He had agreed to it by accepting the job. Was he willing to negotiate for a change in the contract now, in this situation? Was he willing to take it before the stockholders and the employees?

On the other hand, Tom realized that Freddie was not a rotten boy. Everyone liked him. Maybe Freddie was a kid who needed a little more time and a little more support in order to straighten out. Tom had been hoping that a portion of tolerance on his part temporarily might bring about a fundamental change in the boy. Turning a young person around was an investment that his business, as well as the society at large, should support. It wasn’t just that Freddie was his sister’s child; Tom would be willing to make a similar investment in any young person he thought he could help change for the better.

The cat jumped in his lap and settled in for a snooze. Her purring soothed him. As his body relaxed a bit, Tom continued to consider the options. Toward morning he was able to make a decision with which he could live.
HYPOTHESIS 1

DIT P score - Age comparison

<table>
<thead>
<tr>
<th>DIT Principled Score</th>
<th>8th</th>
<th>College</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>27.27</td>
<td>38.39</td>
</tr>
</tbody>
</table>

\[
t = -4.98 \\
p < .0001
\]
HYPOTHESIS 4

Total Event Memory Recall - Age Comparison

Mean % Recalled

8th grade: 34.36
College: 49.43

$t = -.675$
$p < .0001$
HYPOTHESIS 5

Principled Moral Argument Recall

2-way ANOVA
Main effects both significant
Age \( p < .000 \)
P score group \( p < .0002 \)