

Text Processing of Domain-Related Information for Individuals with High and Low Domain Knowledge

GEORGE J. SPILICH, GREGG T. VESONDER, HARRY L. CHIESI, AND JAMES F. VOSS

University of Pittsburgh

This paper is concerned with how previously acquired knowledge affects the processing of new domain-related information. Text processing was studied in two groups differing in knowledge of a particular domain (baseball). A knowledge structure for the domain was constructed, and the text propositions were classified in terms of the knowledge structure. After hearing the text of a half-inning account of a fictitious baseball game, subjects summarized its contents. A completion test followed. Analyses indicated quantitative and qualitative differences in recall for high- and low-knowledge individuals. The results were explained primarily in terms of the greater ability of high-knowledge individuals to relate the actions of the game to the goal structure and to maintain the most important information in a working memory system. Application of the Kintsch and van Dijk (*Psychological Review*, 1978, 85, 363-394) model supported this interpretation and suggested the operation of a complex working memory mechanism which processes both macrostructure and microstructure information. Such a mechanism, it was concluded, is difficult to explain in terms of a typical limited-capacity, working memory buffer. Instead, the results suggest the possible existence of two interacting working memory systems, each of which is concerned with the one type of information, or the operation of a more traditional working memory system combined with an "active" list of macrostructure information that is maintained in long-term memory.

Recent research on discourse comprehension has been concerned with the description of generalized, higher-level knowledge structures and the study of how such structures influence comprehension. Examples of this work include the study of story grammars, which are assumed to provide individuals with a basis for encoding and retrieving text information (Mandler, 1978; Mandler &

Johnson, 1977; Rumelhart, 1975; Thorndyke, 1977) and the study of macrostructures, which are assumed to provide individuals within a given culture (Kintsch & Greene, 1978) with knowledge of a higher-level text structure (Kintsch, 1977; Kintsch & van Dijk, 1975). A similar line of investigation has been concerned with how higher-level knowledge structures, termed schemata (e.g., Anderson, Reynolds, Schallert, & Goetz, 1977) and scripts (Schank & Abelson, 1977) influence text understanding.

This paper extends the research on the influence of knowledge upon the understanding and memory of text. More specifically, a knowledge structure was constructed for a particular domain (baseball).¹ The text used in the study, a half-inning account of a fictitious

The authors wish to thank Jim Pellegrino for his comments on an earlier draft of this paper. The authors also wish to express special thanks to Charles Perfetti for his help in scoring the protocols of the present study. The research reported in this paper was supported by the Learning Research and Development Center, supported in part as a research and development center by funds from the National Institute of Education (NIE), United States Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of NIE, and no official endorsement should be inferred. Requests for reprints should be sent to Dr. James F. Voss, Learning Research and Development Center, 3939 O'Hara Street, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

¹The reasons for using baseball as the knowledge domain of inquiry are considered elsewhere (Chiesi Spilich, & Voss 1979).

baseball game, was analyzed in terms of its propositional structure, and the text propositions were classified in terms of the knowledge structure. The text was presented to individuals of either high baseball knowledge (HK) or low baseball knowledge (LK). The individuals subsequently recalled the text contents and then answered a set of questions based upon the text. The protocols were scored in terms of quantity and quality of recall, with comparisons being made between HK and LK performance. In the next section, a description of a knowledge structure of baseball is presented, especially as related to the goal structure and actions of a baseball game. Subsequently, the question of how

differences in knowledge may influence text processing is considered.

Conceptual Framework

Setting. Setting information is defined as background information that is relevant to the play of a baseball game. Three types of such information are delineated and presented in Table 1. The first is termed *general*, and includes the names of the teams playing, which team is at bat and which is in the field, the inning, and other factors such as weather conditions that are relevant (or potentially relevant) to the play of the game.

The second type is *specific*, and involves information pertaining to a particular team or

TABLE 1
BASEBALL KNOWLEDGE STRUCTURE

Setting				
General:		Teams playing, team at bat, team in field, inning, miscellaneous conditions		
Specific:		Relevant: teams' records as related to goal structure, players' records as related to goal structure		
Enabling:		Irrelevant: team attributes, player attributes Batter at bat and pitcher ready to pitch		
Goal structure				
Team at Bat	Level	Variables	Values	Team in field
Winning game	1	Game outcome	Win-lose	Winning game
Scoring runs	2	Score	Domain of game scores	Preventing runs from scoring
Getting runners on base and advancing runners	3	Pattern of base runners	Eight possible patterns	Preventing runners from getting on base or advancing by making outs
		Outs	0, 1, 2, 3	
Having "Balls," Avoiding "Strikes"	4	"Balls" "Strikes"	0, 1, 2, 3, 4 0, 1, 2, 3	Getting "Strikes," Avoiding "Balls"
Non game actions				
Relevant n o n g a m e a c t i o n s				
Irrelevant n o n g a m e a c t i o n s				

player. There are two types of specific setting information, termed *relevant* and *irrelevant*. The former represents information that is not directly related to the goal structure of a particular game but is related to the performance of a team or player with respect to a goal structure variable, for example, the team's won-lost record, the players' batting averages, and the won-lost records of the pitchers. The latter type of information consists of team or player information that is irrelevant to the goal structure of the game, for example, color of uniforms, the numbers worn by players, and so forth.

The third type of setting information is termed *enabling* information. This consists of information or events usually necessary to initiate the game activity, for example, the members of one team being in the field with the pitcher ready to pitch and the appropriate player of the other team coming to bat.

Goal structure. We assume that baseball has a well-defined, hierarchical goal structure. The highest goal of both teams is winning the game. The second-level goal for the team at bat is to score runs, while for the team in the field it is preventing runs from being scored. The third level for the team at bat consists of getting runners on base or advancing runners while for the team in the field it consists of making the batter or runner(s) out. The fourth level goal for the batter is to increase the number of "Balls" in the ball-strike count, while for the team in the field it is to increase the number of "Strikes." Finally, each level of the goal structure contains at least one variable which can take on a finite number of values. The variables of the goal structure and the values of possible variables are presented in Table 1.

Second, we assume that a baseball game may be defined in terms of game states and game actions. A *game state* is defined by the values of the goal structure variables that exist at any given point during a game. For example, there are two outs, runners are on first base and second base, etc. An account of a

baseball game thus is assumed to consist in part of a series of game states. (Setting information is also generally involved in defining a game state.)

A *game action* is defined as any event or sequence of events of the game that produces a change in the value of one or more of the goal structure variables. The list of game actions is too extensive to present, but examples include the pitcher throwing a pitch that is called a "ball" (change in variable of fourth level), a runner stealing a base (change in variable of third level), the making of an out, a batter getting a hit, a play producing a change in the score, and so on.

Three aspects of game actions should be noted. First, when a game is in a particular state, only a particular set of game actions may occur. (A stolen base can only occur if there is a runner on base.) Second, a new game state is produced when a game action has taken place, for example, when a runner advances a base, a new parameter value of a goal structure variable is assigned, and by definition this constitutes a new game state. Third, because of their relation to the goal structure, game actions are structured hierarchically, with actions producing a change in a variable at a higher level of the goal structure being more important than an action producing a change at a lower level.

In addition to the four levels of actions related to the goals of the game, it is necessary to postulate *auxiliary* actions. These are defined as actions that occur as part of a given play that lead to a game action. To illustrate, consider "The batter hits a ground ball to the third baseman. The ball is picked up and thrown to first base. The batter is out." "The batter is out" is a third-level action, but the hitting of the ground ball to the third baseman and the throwing of the ball to the first base are actions of the game that do not, per se, involve the making of an out, even though they lead to the out and are, in a sense, causative. Actions such as those involved in the play of the game which at their conclusion

result in an action that produces a change in the game state are termed auxiliary actions.²

Finally, it is necessary to acknowledge a category of actions that are not related to producing changes in goal structure variables, termed *nongame actions*. These are actions taking place during a baseball game that do not involve plays that are related to changes in the values of the goal structure variables. There are two types of nongame actions, termed *relevant* and *irrelevant*. The former consists of actions occurring during the game, such as the catcher returning the ball to the pitcher, and so forth. The latter type consists of activity that occurs in the game but is irrelevant, such as the crowd yelling when a particular batter comes to bat.

Processing of Game Information in High-Knowledge and Low-Knowledge Individuals

The primary assumption regarding the processing of game information is that it consists of a matching procedure by which the input information is matched to the individual's knowledge structure. (See Chiesi et al., 1979, for a more complete description.) This assumption leads to the hypothesis that HK individuals should be able to process information more readily than LK individuals because they presumably have a more highly developed knowledge structure.

Within the present framework, the knowledge structures of HK and LK individuals may differ in at least three ways. First, the HK person has a better knowledge of the goal structure of the game, including the variables and their values. Second, the HK individual has a more extensive knowledge of game actions and, more importantly, of how these actions are related to the goal structure. More specifically, for any given game state the HK individual has a better knowledge of what game actions may occur, and how such game

actions are related to changes in the parameter values of the goal structure variables. Similarly, when a given action occurs, the HK person knows more about how this action may produce a change in the game state. Third, the HK individual has a better knowledge of the sequences of actions and state changes that are related to the goal structure, and this possible difference requires special consideration.

The processing of the events of a baseball game is assumed to be a function of the previously mentioned components of one's knowledge structure. The "theme" of a baseball game essentially consists of the actions of the game that are related to winning the game, and these actions vary in importance. Performance of HK individuals would therefore be expected to reflect their knowledge of how the actions are related to the goal structure. More salient actions of the game, for example, should be better remembered than less salient actions. Furthermore, because the baseball game consists of a sequence of game actions, it is necessary for the individual to monitor the game states, that is, the score, positions of runners, number of outs, and so on. Indeed, the game may be viewed as a series of "episodes", and what happened in a previous "episode" is often important to the significance of what is happening or will happen. (During a radio broadcast, the announcer often provides information involving these variables and at a baseball game, the scoreboard acts as an external memory system for monitoring much of this information.) We assume that HK individuals are able to monitor such information more effectively than LK individuals, primarily through the operation of a working memory system that keeps track of the values of the goal-related variables (cf. Chiesi et al., 1979) and that HK individuals are more adept than LK individuals in such monitoring. In other words, the more developed macrostructure of the HK individual enables the HK person to maintain the most salient information in the working memory

² There is a question about where to draw the line regarding which actions are auxiliary and the decision must, to some extent, be arbitrary.

system and to relate the input information to what is being carried in the memory system.

A number of hypotheses based upon the above foundations could be advanced regarding differential text processing in HK and LK individuals, but, in general, HK recall should be superior to LK recall, both in terms of the amount recalled and the nature of what is recalled, that is, HK protocols should contain more goal-related information and should provide a more reasonable account of the sequence of game events.

METHOD

Materials. The half-inning account we used may be found in the Appendix. Readers not acquainted with baseball may note the following. The initial few sentences primarily contain setting information. Subsequently, statements pertaining to each goal level may be found within the passage, as well as statements at each action level (with the exception of the highest level). Also, statements involving baseball information and action statements not directly involved in the account of the game are also found in the text. Finally, the text provides little monitoring information, that is, the score is not repeated, when an out is made it is not stated whether it is the first, second, or third out, the bases occupied by runners are not repeated, and so on.

The text was analyzed in terms of its propositional structure. The list of propositions obtained was the result not only of the text per se but of inferences and statements that could be made at recall. Each proposition was then classified in terms of the goal structure. Table 2 presents a summary of the number and proportion of propositions in the various categories of the goal structure. (A complete description of the propositional analysis is available upon request.)

Procedure. The half-inning account was presented via tape. The subjects were initially informed that they would hear an account of a

TABLE 2
SUMMARY OF PROPOSITIONAL STRUCTURE OF GAME ACCOUNT

Category	Total propositions	Percentage of total
Setting	81	31
General	11	4
Specific-relevant	31	12
Specific-irrelevant	33	13
Enabling	6	2
Goal state	36	14
Level 1	3	1
Level 2	5	2
Level 3	8	3
Level 4	20	8
Game action	98	36
Level 1	0	0
Level 1 (auxiliary)	0	0
Level 2	1	(.4)
Level 2 (auxiliary)	6	2
Level 3	9	3
Level 3 (auxiliary)	26	10
Level 4	8	3
Level 4 (auxiliary)	48	18
Nongame action	49	19
Relevant	13	5
Irrelevant	36	14

part of a fictional baseball game which they would be tested on later. The presentation of the half-inning took 5 minutes and 5 seconds. Following presentation of the materials, there were three testing steps. First, subjects were asked to summarize the account in a one- or two-sentence summary. The summary was written on an 8.5×11-in. response form. In order to insure that the summaries would be brief, two horizontal lines exactly 2 in. apart were marked and subjects were instructed to write the summary between the two lines. All subjects finished the task within 5 minutes. Second, subjects were instructed to write down as much of the entire account as they were able to. Fifteen minutes was allotted for this task, and all subjects finished within the time limit. No subject accepted the offer of a time extension. Finally, subjects were given a booklet which contained 40 completion questions. The questions were related to specific

information contained in the account. One redundant question was thrown out. Subjects were allowed 20 minutes for this task and all subjects finished within the time period. Again, no subject accepted the offer of additional time.

Subjects. Forty-six subjects participated in this study, with 23 designated as high in baseball knowledge (HK) and 23 designated as having low baseball knowledge (LK). (See Chiesi et al., 1979, for a more complete description of the selection procedure.) All HK individuals were male;³ in the LK group, 7 were male and 16 were female. All subjects were tested on both the Davis Reading Test (Form 1A) and a completion test of 45 items of baseball knowledge which included knowledge of the terminology, rules, and procedures of baseball. The mean Davis score for the HK group was 22.30, $SD=4.53$, and the mean for the LK individuals was 21.79, $SD=5.44$. The mean score on the baseball test was 41.83, $SD=1.27$, for the HK subjects and 19.08, $SD=6.93$, for the LK subjects. The LK subjects were thus a "Low-Knowledge" and not a "No-Knowledge" group.

RESULTS

*Protocol Data*⁴

Eight randomly selected HK protocols and eight randomly selected LK protocols were scored blindly by two high-knowledge individuals. The scoring system consisted of giving credit for those propositions correctly recalled from the text structure. The scoring reliability for total correct propositions recalled per subject was .94. The percentage

³ Issues pertaining to subject selection are considered elsewhere (Chiesi et al., 1979).

⁴ The brief summary was scored on a 5-point scale, where 1 = poor, and 5 = excellent. In general, performance was poor, with a mean HK rating of 2.48 and mean LK rating of 1.89, $t(44)=2.26$, $p<.025$. Since these data are considered as secondary to the more extensive summaries, they are not considered further.

agreement on the particular propositions recalled was 81% for the HK protocols and 78% for the LK protocols. The remaining protocols were scored by one individual. In addition to correctly recalled propositions, errors were noted and classified.

The HK individuals recalled more propositions, $M=48.4$, $SD=18.8$, than the LK individuals, $M=30.6$, $SD=12.2$, $F(1,44)=14.53$, $p<.001$, $MS_e=251.57$.

To analyze specific propositional differences between HK and LK recall, the proportion of subjects recalling each proposition was determined. From these, data recall protocols were generated which consisted of propositions recalled by 75 and 50% of the subjects in each of the two groups. These data are presented in Table 3.

The 75% recall text for the LK group merely indicated that the weather was bad and the first batter came to bat. The 75% recall text of the HK subjects provided a much more thorough description of the text although there are two omissions of importance to the goal structure, the wild pitch that advanced runners from first and second base to second and third base and the strikeout that ended the inning.

The 50% recall for the LK group provided a much better account than the 75% LK recall. Indeed, the account resembles the 75% account of the HK individuals, although it also omits the two factors mentioned above as well as the scoring of the run. The 50% account of the HK individuals provides a reasonable and inclusive summary of major events of the half-inning with, however, the final out still omitted. A possible reason for this omission is discussed later.

Another way to consider quality of the recall is to examine the propositions recalled in terms of their position in the goal structure. The proportion of propositions recalled for each of the levels shown in Table 1 is presented in Table 4. Also presented are the results of a series of statistical tests comparing recall for the HK and LK groups.

TABLE 3
SYNTHESIZED PROTOCOLS BASED UPON 75 AND 50% HK AND LK RECALL

High knowledge	Low knowledge
<p>75%</p> <p>The Center City Cougars were playing. The first batter came to bat. The pitcher was in the field. The batter hit a fly ball that was caught and the batter was out. The next batter came up. The batter hit the ball and was safe on second with a double. The next batter came up and hit the ball. He was safe at first base. The batter hit the pitch, the runner on third scored, and the runner on second was out at home plate. The umpires called the game because of rain.</p> <p>50%</p> <p>The Ridgeville Robins were playing the Center City Cougars. The score was 5-3. The Robins were ahead. It was the last half of the fifth inning. The sky was dark and the rain was getting heavier. The first batter came up to bat. The pitcher in the field was left-handed. The pitcher's E.R.A. was 6.00. The batter hit the ball into center field, but it was caught and the batter was out. The next batter came up. This batter lead the league in home runs. The batter hit the ball to left field and arrived safely at second with a double. The next batter came up, hit a ground ball toward the shortstop, and arrived safely at first base. The runner on second stayed on base. The runner from second advanced to third base and the runner on first went to second. The next batter came up and got a hit. The runner on third scored and the runner on second was out at the plate. The umpires called the game because of rain.</p>	<p>The sky was dark and the rain was becoming heavier. The first batter came to bat.</p> <p>The Ridgeville Robins were playing the Center City Cougars. The score was 5-3. The sky was dark and the rain was becoming heavier. The first batter, Jones, came to bat. The pitcher was ready to pitch. Jones flew out. The next batter came up, hit the ball, and made it safely to second base. The next batter came up and made it safely to first base. A runner was out at home plate. The umpires called the game because of rain.</p>

The number of setting propositions recalled was significantly greater for the HK group. Within this category HK recall was significantly greater for general information and for relevant specific information, but not for irrelevant specific information. In effect, this means that HK recall was superior to LK with respect to the teams, the inning, batting averages, the record of the pitcher, and so forth. The significant difference in the enabling setting information indicated that HK individuals acknowledged more batters coming to bat than the LK individuals.

A significantly greater number of goal state propositions was recalled by HK individuals. The analyses of the different types of goal state propositions revealed that this superiority was due to better recall of batters getting on base

and runners advancing as well as outs being made. Interestingly, HK and LK groups did not differ significantly in the statement of which team won, the score, and ball-strike information. The fact that HK and LK recall did not differ significantly for the two highest levels of the goal structure supports our contention that the LK subjects had a reasonable knowledge of the goal structure. The failure to find a ball-strike recall difference reflects the fact that individuals of both groups seldom recalled this type of information. (Ball-Strike information becomes irrelevant once the batter has finished his or her time at bat.)

With respect to game action information, HK recall was significantly greater than LK recall. The within-category analyses revealed

ositions re-
col and 78%
aining proto-
al. In addition
is, errors were

more propo-
the LK indi-
(1,44) = 14.53,

onal differen-
e proportion
position was
call protocols
d of propo-
f the subjects
ese data are

group merely
bad and the
recall text of
much more
xt although
rtance to the
at advanced
se to second
t that ended

p provided a
% LK recall.
s the 75%
although it
ned above as
50% account
a reasonable
vents of the
nal out still
s omission

ality of the
ons recalled
al structure.
recalled for
is presented
: results of a
g recall for

TABLE 4
PROPORTION OF PROPOSITIONS RECALLED AT EACH STRUCTURE LEVEL FOR HIGH AND LOW KNOWLEDGE GROUPS

Category	High knowledge		Low knowledge		Statistical comparison	
	Mean recalled	Proportion	Mean recalled	Proportion	F(44)	MS _e
Setting	19.26	.24	12.56	.16	17.577***	29.366
General	6.09	.55	4.78	.43	4.225*	4.630
Specific (Relevant)	5.52	.18	2.00	.06	25.748***	5.540
Specific (Irrelevant)	3.52	.11	2.78	.08	1.384	4.538
Enabling	4.13	.69	3.00	.50	6.982*	2.105
Goal State	3.26	.09	2.14	.06	3.457	4.251
Level 1	.35	.35	.04	.01	3.307	0.322
Level 2	1.39	.28	.96	.19	2.059	1.055
Level 3	1.09	.14	.57	.07	4.673*	0.669
Level 4	.43	.02	.57	.03	<1	0.802
Game Action	22.73	.23	11.54	.12	18.054***	79.532
Level 2	1.00	1.00	.43	.43	11.841**	0.310
Level 2A	1.48	.25	.56	.09	14.352***	0.668
Level 3	6.17	.69	3.91	.43	10.726**	5.480
Level 3A	9.26	.36	4.43	.17	21.503***	12.456
Level 4	.52	.06	.30	.04	<1	0.832
Level 4A	4.30	.09	1.91	.04	4.692*	14.016
Nongame Action	3.13	.06	4.30	.09	1.442	10.988
Relevant	.39	.03	.52	.04	<1	0.437
Irrelevant	2.74	.08	3.78	.11	1.397	8.962

* $p < .05$.

** $p < .01$.

*** $p < .001$.

substantial differences in the recall of the scoring play of the inning, as well as outs being made and batters getting on base or runners advancing. Interestingly, the HK group also recalled a greater number of auxiliary propositions that were parts of actions producing changes in the variables of the goal structure. This finding thus indicates that HK individuals recalled more information than LK individuals pertaining to how game actions occurred. There were no significant differences in the recall of actions related to balls and strikes, although HK individuals recalled more information about how "Balls" or

"Strikes" occurred, for example, the type of pitch the batter missed when he struck out. The HK and LK individuals did not differ significantly in their recall of either relevant or irrelevant nongame actions although LK recall was greater in both cases.

Taken as a whole, the results summarized in Table 4 indicate that while LK individuals apparently had some knowledge of the goal structure of baseball, the HK protocols contained more information pertaining to actions that produced changes in the more significant aspects of the game, that is, scoring, actions of batters and runners, and outs. The HK proto-

cols also contained more information than LK protocols regarding background information about the players, for example, batting averages. (There were some interesting differences of HK and LK performance with respect to the recall of specific propositions, but these are technical, and thus may be more meaningful to baseball fans.)

Confusions. One indicator of processing differences in HK and LK individuals is the nature of confusion errors. The primary confusion of HK individuals was inappropriate assignment of batting averages; typically the batting average recalled was stated in the text but attributed to the wrong person. An average was stated for four of the five batters mentioned in the text, and there were 39 statements of averages in HK protocols. Nineteen of these were incorrect. For the LK individuals, 10 averages were stated and 5 were incorrect. Thus, the percentage correct was equivalent, but the frequencies were different. This finding is of interest because it implies that HK individuals apparently viewed batting averages to be of some importance, but they had trouble in forming the appropriate player-average associations. The LK individuals, on the other hand, had the same trouble, even though their frequency of recall was lower. For comparison purposes, we note that 10 HK individuals recalled the pitcher's earned-run average correctly, with no errors, while 3 LK individuals recalled it, also with no errors. This finding thus suggests that the relatively high error rate for both HK and LK individuals for batting averages was essentially an associative interference problem.

Second, and quite importantly, LK individuals made 33 player-action errors whereas HK individuals made only 4 such errors. This type of error involves two types of confusion that cannot be separated, namely, attributing a particular action to a player other than the one who performed the action and confusing the order of text events. The LK errors were usually not simple, that is, they

were not of the type "The first batter doubled" when in fact the second batter did. The errors attributed the actions of two players to one batter, and actions recalled were, in a few instances, not stated in a reasonable order. Indeed, on occasion, LK accounts were difficult to follow. The confusion results thus strongly indicate that LK individuals had much more trouble keeping track of the order of the game events. By comparison, the HK individuals not only made fewer such errors but also almost invariably stated information in the appropriate order. Thus, although a percentage-correct order measure may not be derived because of problems in scoring, it nevertheless is quite clear that HK individuals were highly superior in this regard. Interestingly, HK and LK individuals made about the same number of errors in confusing the order of the batters, 10 and 11, respectively, when they used the batters' names, for example, "The next man up, Johnson,..." This finding at least suggests that HK and LK difference-in-order confusions were due primarily to LK individuals having much more trouble keeping track of the order of actions, not the order of players.

Most of the other types of confusion were of relatively low frequency and roughly equivalent in the HK and LK groups, for example, wrong inning (6-6), wrong team ahead (4-3), wrong field to which a ball was hit (9-9), respectively.

General findings. Before considering the results of the completion test, four general aspects of the protocol data are considered. First, HK individuals occasionally elaborated events and made them more graphic, for example, the pitcher was referred to by one person as the "big, fastballing lefthander." The text stated that the pitcher was left-handed and typically struck out quite a few batters. The subject's statement was apparently based upon the inference that pitchers who strike out quite a few batters are most often fastball pitchers and such pitchers also tend to be of large stature, even though there are excep-

LEDGE GROUPS

atistical
comparison

MS_e

29.366
4.630

5.540

4.538
2.105

4.251

0.322

1.055

0.669

0.802

79.532

0.310

0.668

5.480

12.456

0.832

14.016

10.988

0.437

8.962

; the type of
e struck out.
lid not differ
er relevant or
lthough LK

mmarized in
L individuals
e of the goal
rotocols con-
ing to actions
re significant
ng, actions of
ie HK proto-

tions. Second, the HK subjects were much more likely to report larger chunks of information for a particular action, for example, the LK subject stated that the particular batter "got a double," whereas the HK subject was more likely to say that the "batter lined a double down the left-field line." The latter statement is not only accurate in terms of the text but it involves the tendency of HK subjects to unitize a given sequence of actions. Indeed, on occasion the HK subjects went farther than the text, for example, "... hit a line drive against the left field wall," or "hit a foul ball to the right that went into the stands." Third, as indicated by the confusion data, HK protocols provided a much more accurate account of the order of information when such information was related to the goal structure. While the HK subjects occasionally omitted significant information, the information stated was in appropriate order; the LK subjects both omitted more information and stated information in an inappropriate order more often. Fourth, it was somewhat of a surprise that HK individuals often omitted the third out, an important point in the account. This out not only ended the inning but made the game official. The explanation for this finding rests upon the monitoring process discussed earlier. The account contained no updating statements of the number of outs that had been made, and we think that some subjects lost track of the number of outs. As noted previously, play-by-play accounts of games given through the media, especially on radio, are marked by redundant statements that keep listeners updated with respect to the values of the goal structure variables. Indeed, two of the HK subjects did their own updating. For example, one wrote, "Runners now on first and second and one out."

Of necessity, the discussion of the protocol results involved direct reference to the contents of the protocols. For the reader who is not a baseball fan this discussion may seem demanding and/or unclear. For this reason we shall now state a description of the protocol

findings in general terms. The HK subjects recalled more information that was goal related than the LK subjects, they tended to integrate sequences of goal-related actions more than LK individuals, and they recalled information in the appropriate order more than LK individuals. Setting information of possible importance to the goal structure was generally recalled more frequently by the HK individuals, but information and actions not related to the goal structure were recalled more frequently by LK individuals. If one were to characterize the HK and LK differences in general terms, LK recall was like that of a free recall list, excluding a recency effect, in which primacy information was recalled and other specific items were also recalled. There was some organization in LK protocols, but the omissions often showed large gaps in the organization and the confusions suggested LK subjects had considerable trouble organizing information sequentially. For HK subjects recall was more coherent, both with respect to the sequential nature of the account and to the completeness of description of important events. The HK individuals did, however, have some trouble recalling events late in the account, and this difficulty is attributed to problems of monitoring.

Questionnaire results. As expected, the HK group, $M=21.50$, $SD=4.75$, gave more correct answers than the LK group, $M=13.00$, $SD=4.48$, $F(1,44)=39.0$, $p<.001$, $MS_e=21.31$. Of greater interest, however, are the qualitative differences in HK and LK recall.

First, the results indicated that both HK and LK subjects knew more of the account than they stated in their protocols, although this finding was more prevalent for HK individuals. For example, all HK individuals answered a question about weather conditions correctly, but only 15 referred to the weather in the protocols. Also, 20 HK subjects stated the earned-run average of the pitcher, although only 10 of these subjects stated it in

s. The HK subjects in that was goal re-acts, they tended to goal-related actions ls, and they recalled appropriate order more ting information of e goal structure was equently by the HK on and actions not ture were recalled individuals. If one HK and LK dif-LK recall was like xcluding a recency nformation was re-items were also rganization in LK ons often showed tion and the con-jects had con-izing information cts recall was more t to the sequential o the completeness t events. The HK ave some trouble he account, and l to problems of

expected, the HK 5, gave more cor-group, $M = 13.00$, 9.0 , $p < .001$, rest, however, are in HK and LK

ed that both HK e of the account otocols, although alent for HK in-HK individuals it weather con-5 referred to the), 20 HK subjects ge of the pitcher, jects stated it in

the protocol. Eight of the LK subjects stated it correctly and only 3 of these stated it in the protocol. These results thus suggest that recall, especially for HK individuals, was structured according to the goal structure and the related actions, and that lower-order information either was not retrieved or edited and not stated (cf. Mandler, 1978).

Second, one question asked the batting average of the second batter (the 40 questions followed the general account of the game), and this batting average was not in the text. Nineteen HK individuals generated an average, usually a figure in the text that was stated for a different batter. However, 2 HK (and no LK) individuals stated explicitly that the batting average requested had not been stated in the text. This result supports the notion of a schema-like structure which permits the assignment of particular parameters to the goal structure variables.

Third, HK individuals recalled much more about actions of particular batters than LK individuals, for example 20 HK individuals correctly answered that the first batter flew out to center field, whereas 9 LK individuals recalled the first batter made an out.

Fourth, the 23 HK subjects recalled a total of 9 names of players in the field (excluding the batter), whereas LK individuals recalled a total of 25 names. The names of 8 of the 9 players in the field had been stated in the account. This result depicts the tendency of LK individuals to recall more irrelevant information than HK individuals.

Finally, in general, HK subjects gave more correct answers to questions about major game actions than LK subjects. Probably the most interesting example of this difference is that 18 HK individuals correctly described the scoring play in which one runner scored and the second runner was thrown out at home plate. Ten LK subjects gave the correct answer.

DISCUSSION

The first issue considered is the relation of HK and LK performance differences to the

conceptual framework presented in the introduction. Within this framework, HK and LK differences could be found (a) in knowledge of the goal structure, including knowledge of the goal structure variables and their parameter values; (b) in knowledge of how the actions of the game are related to the goal structure; and (c) in the processing of the sequences of game actions. The scores on the initial baseball test as well as the results shown in Table 3 suggest that, at least in the present case, LK individuals had a reasonable knowledge of the goal structure. We assume therefore that the explanation for HK and LK performance differences may be found in the second and third factors.

The results of other work (Chiesi et al., 1979) indicated that presentation of contextual information at input enhanced subsequent recall for the individuals even when the contextual information was not present at recall. However LK subjects showed poorer recall of target information with increasing contextual information present at input. In a sense the present findings amplify this result. While it may be reasonably assumed that LK individuals know that getting a hit is important, they nevertheless may have trouble in following consequences of the hit that are related to the goal structure, for example, following the advancement of base runners. Furthermore, if it is difficult for LK individuals to process the consequences of an event such as a hit, then processing of longer sequences of events should be all the more difficult, especially since longer sequences of significant events are typically interspersed with less significant events. The LK subjects, in other words, have difficulty in monitoring the values and changes in values of the goal structure variables. Indeed, as previously noted, HK individuals are similarly, but less seriously, affected when information indicating the values of the goal structure variables is not repeated. While the data support the general interpretation just outlined, we endeavored to provide a more quan-

titative evaluation of these ideas by utilizing the model of text processing published by Kintsch and van Dijk (1978). These authors assume that a text is processed in terms of processing units which they call cycles. While what constitutes a cycle is somewhat arbitrary, they defined cycles in terms of propositions existing in a sentence. We however defined cycles on the basis of game episodes because baseball episodes seemed intuitively to be more suitable in the game context. Within each cycle Kintsch and van Dijk construct a coherence graph, a branching structure which provides an organization of the propositions within each cycle. Similarly, we constructed a coherence graph, with two individuals independently constructing both the cycles and coherence graph of our half-inning passage. The results were in substantial agreement, with any differences being resolved by selecting the account most closely related to the logic of the game. Finally, Kintsch and van Dijk delineate two types of propositions, macrostructure and microstructure. The latter are arrived at via propositional analysis of the text while the former involve propositions of the text or derived from the text that are related to a hypothetical macrostructure. In our description, we also delineated these two types of propositions, with macrostructure propositions being specifically defined in terms of the game's goal structure. (We were also able to establish a classification of macroproposition in terms of the relation of the macropropositions to the knowledge structure.)

Having established the graphic structure of the text, the next step was to develop and test appropriate models of processing for HK and LK individuals. The Kintsch and van Dijk model involves four parameters, s , p , m , and g . (For our purposes the role of g is minor and g is therefore not considered.) The s parameter denotes an assumed working memory buffer which carries s propositions from one cycle to the next. The model assumes that the propositions carried over are determined by

importance (level in the cycle structure) and recency. The p parameter denotes the probability of recall of micropropositions. The m parameter refers to the probability of recall of macrostructure propositions. Upon developing a specific model, sets of equations are generated which relate the p and m parameters to the recall of particular propositions. The s parameter is taken into account by the propositions that are assumed to be carried over from cycle to cycle, and the assumption is made that for each new cycle in which a given proposition occurs there is an increase in the probability of that proposition getting into long-term memory and being recalled. Given the generated equations, estimates of the parameter values may be obtained by the STEPIT procedure (Chandler, 1965) using the frequency of recall for each proposition for estimating purposes.

Our first series of analyses was conducted to determine which s value gives the best fit for the HK and for the LK data. A priori we expected that the s value may be greater for HK than for LK individuals because the former may be able to hold more information in working memory. However, using s values of 0, 1, 2, 3, 4, and 6, the best-fitting value for both the HK and LK data was $s=2$. The HK $RMSD=.2236$ and the LK $RMSD=.1578$. That $s=2$ provided the best fit is in agreement with the Kintsch and van Dijk results. As with Kintsch and van Dijk, however, we found the value of s to have a relatively small effect in terms of the fit.⁵ Using the $s=2$ model, we found $p=.164$, $m=.106$ for HK individuals, while $p=.105$ and $m=.087$ for the LK group, that is, the parameter values indicate greater recall by HK individuals for both microstruc-

⁵We generally used a least-squares method in the STEPIT procedure although for this series of models, $s=0, 1, 2$, and so on, we also used a minimum χ^2 procedure with equivalent results. χ^2 tests for all values of s indicated that each model need not be rejected, that is, χ^2 was not significant. However, $s=2$ also gave the best fit using χ^2 .

cle structure) and denotes the probabilities. The m stability of recall of s . Upon development of equations are and m parameters propositions. The s count by the probability to be carried over the assumption is s in which a given increase in the proportion getting into recall. Given estimates of the s obtained by the (1965) using the 1 proposition for

was conducted to see the best fit for data. A priori we may be greater for HK individuals because the more information we have, using s values the best-fitting value for $s=2$. The HK $RMSD=.1578$. It is in agreement with the LK results. As with the LK, we found the very small effect in the $s=2$ model, we found HK individuals, or the LK group, indicate greater both microstructure

uses method in the series of models, minimum χ^2 procedure values of s indicated 1, that is, χ^2 was not the best fit using χ^2 .

ture and macrostructure propositions. To this point the analyses provided no evidence for the notion that HK individuals have some type of "expanded" working memory buffer compared to LK individuals, although HK recall was superior to that of LK recall.

The analysis reported thus far essentially constitutes an application of the Kintsch and van Dijk model. The model was extended in the analyses now considered. The Kintsch and van Dijk model considers macrostructure propositions as having a separate value which is, in a sense, superimposed upon the micropropositional structure. In the present work, however, it would be expected that HK and LK differences would occur in terms of the processing of macrostructure propositions. We therefore developed a number of models involving macrostructure operations.

The first model considered is as follows. A macropropositional buffer was assumed which carried propositions of importance to the logic of the game. If, for example, an out was made, then the macroproposition indicating that out was carried over until the next out was made. Similarly, for a given batter, the particular values of balls and strikes would be maintained, dropping the old ball-strike count with each successive ball or strike until the batter either made an out or got on base. Then the number of balls and strikes on that batter would be dropped. We developed a set of equations involving the carry-over of macrostructure propositions from cycle to cycle according to "baseball logic," such as that previously outlined. Included in this logic was also the general setting information carried over as appropriate. In addition, we developed a new parameter, P , which was analogous to p , that is, P was based upon propositional carry-over from cycle to cycle, but only for macropropositions. According to this model, for HK individuals, $P=.023$, $p=.149$, $m=.180$, and $RMSD=.2290$. For LK individuals, $P=.017$, $p=.089$, $m=.111$, and $RMSD=.1573$. While these findings were suggestive, the similarity in

$RMSD$ values to the previous analyses indicates this model did not provide a better fit than the model previously described.

Without going into detail, we note that three other models were tried which involved the following assumptions. The first did not assume a complete macrostructure carry-over based upon "baseball logic" as outlined above, but assumed a macrostructure buffer having two slots in which macrostructure propositions would be carried over from cycle to cycle. A second model assumed a macroproposition buffer having four macrostructure propositions instead of two. Both of these models provided poorer fits than either of the two previously described models.

The best-fitting model we obtained assumed the "baseball logic" macrostructure carry-over buffer described above, combined with the assumption of an $s=2$ microproposition buffer. Furthermore, there are two variations of this model which are of interest. In the first, macrostructure propositions are assumed to be in a macrostructure buffer while micropropositions are assumed to be in a microproposition buffer. This will be termed the independence model. In the second version, a separate macrostructure buffer was assumed, but macropropositions were permitted to have microproposition status and thus, when appropriate, could also be carried over in the microstructure buffer as well as in the macrostructure buffer. The second version provided the better fit. For HK individuals, $P=.019$, $p=.146$, and $m=.044$, $RMSD=.2136$. For LK individuals, $P=.014$, $p=.090$, and $m=.028$, $RMSD=.1469$. In this case the P parameter denotes the probability of recall of macropropositions attributed to cycle-to-cycle carry-over.

We were able to refine both versions of the model further by using five additional parameters to estimate values for the propositions at various levels of the goal structure. These results are not presented, but suffice it to say that under such conditions, the fit of the modified second-version model yielded

RMSD values of .1922 and .1348 for the HK and LK data, respectively.

While the *RMSD* values provide some idea of goodness of fit, we were able to obtain correlational data related to the fit of the models. Specifically, using the three parameters of the best-fitting, nonindependence model, the estimated values of the recall probabilities for the propositions at the various levels of the goal structure were correlated with the actual data shown in Table 4. For the HK individuals, the correlation of predicted and actual data was $r(10) = .730$, $p < .01$. None of the other models for HK individuals which involved some type of macroproposition processing structure yielded a significant correlation. For LK individuals, the nonindependence model produced $r(10) = .830$, $p < .001$. (Two other macroproposition structure models produced correlations significant at the .05 level for LK individuals.)

The analyses presented have a number of interesting implications. First, they support the Kintsch and van Dijk model and indicate one way it may be modified to meet a given set of conditions such as having HK and LK individuals. Second, the findings suggest that both HK and LK individuals have a complex working memory mechanism which enables them to hold macrostructure and microstructure information. Because the best-fitting model for both HK and LK individuals involved the carry-over of two micropropositions as well as a relatively large number of macropropositions, an interpretation based upon a finite-capacity short-term memory buffer seems inadequate. This seems to be especially the case when consideration is given to the fact that a number of less significant events often occur between successive macrostructure-related events.

Two alternative interpretations are considered. First, there may be two interacting buffers, one handling macrostructure information and the other handling more local and adjacent microstructure information. The latter buffer is of the nature assumed by

Kintsch and van Dijk and is the typical finite-capacity, working memory structure. The former also may be assumed to have a finite capacity, but it seems to be capable of holding macrostructure information during the processing of microstructure information. The fact that the nonindependence model provided a better fit than the independence model suggests, however, that appropriate macropropositions may also be processed in the microproposition buffer. An alternative interpretation is that there is a finite-capacity working memory system but that macrostructure information is kept on an "active" list in long-term memory and thus may be readily accessed. The present results do not provide a basis for choosing between these alternatives. However, the findings do point to the need to postulate some type of an active system that holds macrostructure information while microstructure information is being processed.

One argument that could be offered in relation to the apparently large macroproposition carry-over capacity is that the macrostructure information is "chunked" better than the microstructure information. Although Chase and Simon (1973) used the "chunking" concept to interpret the functional significance of a spatial arrangement of chess pieces and we are studying the temporal integration of information, we would nevertheless tend to agree that "chunking" may be seen in, for example, the ability of the HK individual to describe the auxiliary action of a play better than the LK individual. However, it is difficult in terms of "baseball logic" to see why individuals would chunk the inning, scores, names of teams, number of outs made thus far in the inning, current position of base runners, and ball-strike count on the batter. These things simply do not go together; the alternative of some type of macrostructure processing mechanism seems more reasonable.

With respect to HK and LK differences, we would mention that one reason why the LK

data yielded better "fits" than HK data for all models is that with the greater HK recall, there also was greater variation in recall and this produced greater difficulty in the fitting procedure. Of greater importance in considering HK and LK performance is the finding that, as expected, the p and m parameter values were greater for HK than LK individuals. This lack of a substantial difference in the P parameter for HK and LK individuals suggests that those macro-structure propositions the LK individuals were able to recall they were able to carry over from cycle to cycle. They simply did not recall as many as HK individuals, as the data of Table 4 show. Finally, the p parameter differences suggest the interesting result that while both HK and LK individuals carry over a similar number of propositions from cycle to cycle, the HK individuals tend to recall more propositions from within a cycle and quite likely carry propositions over from cycle to cycle more systematically than do LK individuals.

The present research is, of course, closely related to notions suggesting that schemata operate on input information and that such processing provides for a particular interpretation of input information (Anderson, 1977). The contribution of the present work to schema theory, at least as we see it, is that the conceptual model presented earlier as well as the utilization of the Kintsch and van Dijk model provide a reasonably specific account of how schemata may be influential in the processing of specific domain-related information. This specificity was described by the explicit statement of the game's goal structure and hypotheses relating how knowledge of the goal structure may influence the processing of the text.

APPENDIX

The Ridgeville Robins are playing the Center City Cougars. The Robins are leading 5-3 with the Cougars at bat in the last half of the fifth inning. The sky is getting darker, and the rain that has started is becoming heavier.

The Cougars' first batter, Harvey Jones, is taking his time coming to plate. The umpire steps back from behind the plate and tells him to step into the batter's box.

Jones, the hitter, is left-handed, and has a batting average of .310. Claresen, the pitcher, has allowed only four hits, has walked one, and has struck out six. This performance is about average for Claresen since this left-hander has an earned-run average of 6.00 and typically strikes out quite a few batters.

Claresen now adjusts his cap, touches his knee, begins his windup, and delivers a high fastball that the umpire calls "Ball One." The Robins' catcher, Don Postman, returns the ball, and Claresen takes the sign. The next pitch is swung on and hit to centerfield. Maloney comes in and catches it for the first out.

The next batter is the powerful hitter, Fred Johnson, who leads the league in home runs with 23. Claresen no doubt is glad to face him with no one on base. Claresen is now getting his sign from the catcher, begins his windup, and throws a curveball breaking into a right-hand batter at the knees. The umpire calls it a strike. Claresen is now getting ready again, winds up and throws, and Johnson hits it off to right and into the stands, a foul ball. The count is now 0 and 2. Claresen rubs up the new ball, takes his sign, and throws a fastball which just misses the bill of Johnson's cap. Johnson took one step toward the mound, but then came back. Johnson stepped out of the batter's box, and put some resin on his hands; the bat is no doubt slippery from the rain. Claresen is ready again, winds up and throws a slider which breaks inside, making the count 2-2. Once more Johnson steps out of the batter's box and gets some resin on his hands. He steps back in and Claresen starts his motion and throws. Johnson swings and has a line drive down the left-field line. Ferraro runs over to get the ball as Johnson rounds first and goes toward second. Ferraro's throw is late and Johnson is safe on second with a double.

Beck, the left-handed relief pitcher, is warming up in the bull pen. The next hitter for the Cougars is the right-hand hitting Carl Churniak, a .260 hitter who is known to hit well in the clutch. Claresen takes his sign, delivers, and Churniak takes the pitch for a ball. Claresen again is ready and pitches, and Churniak swings and hits a slow bouncing ball toward the shortstop. Haley comes in, fields it, and throws to first, but too late. Churniak is on first with a single, Johnson stayed on second.

The next batter is Whitcomb, the Cougars' left-fielder. He is a left-hander hitting .255. Claresen wipes his forehead with his sleeve and takes his sign. Claresen looks toward first, where Manfred is holding the runner. He stretches, looks at second, and throws a high fastball for a strike. The catcher returns the ball and Claresen once more gets ready. Claresen throws a low curveball. It bounces into the dirt and past the catcher. Johnson moves to third and Churniak to second before the catcher can retrieve the ball. The ball is returned to Claresen. He gets

the sign and winds up, and throws a slider that Whitcomb hits between Manfred and Roberts for a hit. Dulaney comes in and picks up the ball. Johnson has scored, and Churniak is heading for the plate. Here comes the throw, and Churniak is out. Churniak argues, but to no avail. The batter reached second on the throw to the plate.

The next batter is Rob Williams, the Cougars' catcher. He is hitting .230. Claresen is rubbing up the ball and now is ready to pitch. The rain is coming down in sheets. Claresen delivers and Williams takes a curve over the inside corner for a strike. Working rapidly Claresen again delivers and Williams takes a ball, low and outside. Claresen again gets the sign, stretches, and throws a fastball, which Williams swings and misses. The catcher returns the ball and Claresen is ready. The pitch is a curveball which Williams swings at and misses for his third strike.

The umpires now are meeting and they signal that the game is being called.

REFERENCES

- ANDERSON, R. C. The notion of schemata and the educational enterprise: General discussion of the conference. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.
- ANDERSON, R. C., REYNOLDS, R. E., SCHALLERT, D. L., & GOETZ, E. T. Frameworks for comprehending discourse. *American Educational Research Journal*, 1977, 14, 367-381.
- CHANDLER, J. P. *STEPIT Program 90PE66, Quantum chemistry program exchange*. Bloomington, Ind.: Indiana University, 1965.
- CHASE, W. G., & SIMON, H. A. Perception in chess. *Cognitive Psychology*, 1973, 4, 55-81.
- CHIESI, H., SPILICH, G. & VOSS, J. F. Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, 1979, 18, 257-273.
- KINTSCH, W. On comprehending stories. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.
- KINTSCH, W., & GREENE, E. The role of culture-specific schemata in the comprehension and recall of stories. *Discourse Processes*, 1978, 1, 1-13.
- KINTSCH, W., & VAN DIJK, T. A. Comment on se rapelle et on résume des histoires. *Langages*, 1975, 40, 98-116.
- KINTSCH, W., & VAN DIJK, T. A. Toward a model of text comprehension and production. *Psychological Review*, 1978, 85, 363-394.
- MANDLER, J. M. A code in the node: The use of a story schema in retrieval. *Discourse Processes*, 1978, 1, 14-35.
- MANDLER, J. M., & JOHNSON, N. S. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 1977, 9, 111-151.
- RUMELHART, D. E. Notes on a schema for stories. In D. Bobrow & A. Collins (Eds.), *Representations and understanding: Studies in cognitive science*. New York: Academic Press, 1975.
- SCHANK, R., & ABELSON, R. *Scripts, plans, goals, and understanding*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.
- THORNDYKE, P. W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, 9, 77-110.

(Received June 30, 1978)