The effects of mental representation on performance in a navigation task

IMMANUEL BARSHI and ALICE F. HEALY
University of Colorado, Boulder, Colorado

In three experiments, we investigated the mental representations employed when instructions were followed that involved navigation in a space displayed as a grid on a computer screen. Performance was affected much more by the number of instructional units than by the number of words per unit. Performance in a three-dimensional space was independent of the number of dimensions along which participants navigated. However, memory for and accuracy in following the instructions were reduced when the task required mentally representing a three-dimensional space, as compared with representing a two-dimensional space, although the words used in the instructions were identical in the two cases. These results demonstrate the interdependence of verbal and spatial memory representations, because individuals' immediate memory for verbal navigation instructions is affected by their mental representation of the space referred to by the instructions.

Our focus in this study was on the mental representations that individuals form when following verbal navigation instructions. We also examined how these representations influence the individuals' ability to repeat and follow the instructions.

To navigate in space, individuals have to form a spatial representation. When the individuals must follow verbal instructions in order to navigate, it may be possible to map the language input directly onto the spatial representation. However, in situations in which it is important to verify that instructions were correctly received prior to their execution, it is common to require that they be repeated back, and in such cases, a verbal representation must be formed in addition to the spatial representation. The nature of the relationship between these two representations, as well as the nature of the representations themselves, have been a source of controversy.

Garden, Cornoldi, and Logie (2002) suggested that both verbal and visuospatial aspects of working memory (WM) contribute to performance on a navigation task.

Shah and Miyake (1996) offered evidence to support the separation of spatial thinking and language processes in WM. Similarly, Toms, Morris, and Foley (1994) supported the separation of visuospatial and verbal WM processes. More recently, Friedman and Miyake (2000) argued that spatial and causal dimensions of mental representations are handled independently in different WM subsystems. These recent claims can be contrasted particularly with the earlier work by McNamara and LeSueur (1989), who showed that nonspatial relations influence the probability of encoding spatial relations, suggesting some interdependence between these two types of memory processing. One goal of the present study was to address this debate concerning the independence or interdependence of the two representation systems, as well as to shed light on the nature of the representations themselves.

We assume that when navigating in space, individuals form a mental model (Johnson-Laird, 1983) of the space. The assumed spatial representation is similar to van Dijk and Kintsch's (1983) *situation model* (see, also, Kintsch, 1988; Zwaan & Radvansky, 1998), in that it is constructed on the basis of the given situation and is used in the interpretation of language input. Of concern in the present study was the situation model used in the context of the following of verbal navigation instructions. A situation model may involve numerous conceptual dimensions. Investigators have focused on different conceptual dimensions in their treatments of situation models. For instance, Rapp, Gerrig, and Prentice (2001) focused on trait-based situation models in their study of inferences made about characters in narratives; Zwaan, Madden, and Whitten (2000) focused on the temporal continuity and discontinuity of events in their study of narrated situations; and Morrow, Greenspan, and Bower (1987; see also Morrow, Bower, & Greenspan, 1989) focused on spatial locations in their studies of readers' ability to assume a protagonist's

The experiments reported here were part of a doctoral dissertation by I.B. under the guidance of A.F.H. This research was supported in part by Contracts MDA903-93-K-0010, DASW01-96-K-0010, and DASW01-99-K-0002 from the Army Research Institute, by Grant DAAH04-95-I-0208 from the Army Research Office, and by Grants NCC2-1112 and NCC2-1310 from the National Aeronautics and Space Administration to the University of Colorado (A.F.H., principal investigator). The authors thank Ernest Mross for help with computer programming, Bob Rehder for recording the instructions, Alison Clark for help with data collection (Experiments 2 and 3), Michael Allen for help with stimulus construction, and Lyle Bourne, David Bryant, Zygmunt Frajzyngier, Michael Kahana, Robert Kail, Walter Kintsch, Thomas Landauer, and Timothy McNamara for helpful comments about this research. Correspondence concerning this article should be addressed to I. Barshi, who is now at the Human Factors Research and Technology Division of the National Aeronautics and Space Administration Ames Research Center, Mail Stop 262-4, NASA Ames Research Center, Moffett Field, CA 94035-1000 (e-mail: ibarshi@mail.arc.nasa.gov).

perspective in constructing their own situation models of narratives. The present study focused only on spatial dimensions, so we considered the relevant situation model to be equivalent to a spatial representation. Specifically, we focused on whether the number of spatial dimensions along which one travels and/or the number of spatial dimensions included in the spatial representation affect individuals' ability to execute navigation instructions.

Whereas spatial representations can be seen as a type of situation model, it has been argued that verbal representations are propositional in nature (e.g., Kintsch, 1972, 1994; Kintsch & Keenan, 1973; Perrig & Kintsch, 1985). Alternatively, the verbal representation could be word based, as in verbal learning experiments (e.g., Murdock, 1962). However, it is likely that different verbal representations would be used in different circumstances. Of concern in the present study was the verbal representation used in a context in which verbal navigation instructions are followed. Specifically, we focused on whether the verbal representation formed, as evidenced by individuals' ability to repeat back the words used in the navigation instructions, would be affected by the spatial representation necessary for executing the navigation task.

We will report three experiments in which a paradigm was used that was developed by Barshi (1997) to investigate navigation in a 3-D space shown as a grid on a computer screen. In Experiment 1, we explored the nature of the verbal representation; in Experiment 2, we explored some aspects of the spatial representation; and in Experiment 3, we explored the relationship between the spatial and the verbal representations, with the aim of determining whether they are independent or interdependent. In all three experiments, we examined immediate verbal repetition of the navigation instructions, as well as execution responses. The oral repetition responses provided a measure of immediate recall and, hence, reflected the verbal representation, and the manual movement responses provided a strict measure of message comprehension and, hence, reflected the spatial representation. Specifically, in Experiment 1, the question was probed of whether the unit involved in the comprehension of the navigational instructions was based on words or propositions. In Experiment 2, we probed the differences between movements along three dimensions and movements along only two dimensions of the same 3-D space. In Experiment 3, we probed the differences between movements on the picture plane (i.e., the plane shown directly on the computer screen) and movements on a plane orthogonal to the picture plane in response to the same directional terms (i.e., right/left and up/down). Of primary interest in Experiment 3 was whether the planes along which participants moved affected their ability to provide an immediate oral repetition of the instructions, even though the words used to instruct movement along the different planes were held constant.

GENERAL METHOD

Because of its critical consequences, Barshi's (1997) paradigm was designed to be analogous to communication between flight

crews and air traffic controllers. Similarly, in the present study, the participants played an analogous role to that of a pilot who is receiving air traffic control (ATC) instructions. The participants heard messages instructing them to move in a simulated 3-D space displayed on a computer screen. They repeated those instructions orally and then followed them. Unlike ATC instructions, the experimental instructions did not include any aviation jargon. The instructions were to turn right or left, climb up or down, and move forward or back (the *forward* and *back* instructions referred to movement within the space—i.e., forward meant away from the participant's position; this orientation is consistent with a spatial framework analysis; Franklin & Tversky, 1990). The wording of the experimental instructions was kept to simple standard English in order to eliminate the need to train the participants in the special terminology of aviation English.

Apparatus and Materials

A male native English speaker was recorded giving different combinations of the instructions in as natural a manner as possible. His voice was digitized on a Macintosh SE30 computer using the program SoundEditPro. The clearest sample of each word used in the instructions (e.g., "one," "up," "forward") was then spliced out of the natural speech stream. A Macintosh II computer played the appropriate words according to a specific sequence indicated for each experimental trial. This process retained a natural stress and intonation pattern within each isolated word, although not over the whole trial. This procedure ensured that the speech rate was held constant across all trials in all three experiments (see Barshi, 1997, for a discussion of the effects of speech rate in this paradigm).

Design

The participants saw four matrices on the computer screen, stacked one above the other (see Figure 1) and were told that the grid represented a 3-D space, which was also illustrated on a small model. Each matrix was made of 16 squares (4×4). The starting position (which was constant) was highlighted. Because there were four 4×4 matrices, the numbers in the instructions were limited to one, two, and three. Also, from any position, there was a limited number of moves the participants could make without "falling off" the grid. The instructions were set so that the participants would never fall off the grid if they followed the instructions correctly. The instructions followed a consistent structure: turn (left, right) (one, two, three) square(s), climb (up, down) (one, two, three) level(s), move (forward, back) (one, two, three) step(s). The instructions were also consistent in terminology. *Turn* was always associated with *right* or *left* and with a number of *squares*; *climb* was always associated with *up* or *down* and with a number of *levels*; and *move* was always associated with *forward* or *back* and with a number of *steps*. The participants were never instructed, for example, to "turn up" or to "move two levels." Such consistency is like that of flight operations, where, for instance, *turn* is always followed by *right* or *left*, the word *heading*, and a compass direction; *climb* is always followed by an altitude; and *contact* is always followed by a facility name and a radio frequency. The experiment simulated the aviation scenario in one more aspect: that of the structure of the message. The instructions followed a fixed order, as do ATC messages. There were three kinds of movements the participants could make—turn, climb, and move—and they were always in that order (i.e., the three kinds of movements were not counterbalanced across message positions). However, because the experimental instructions were limited to movements along three axes, to generate instructions that contained more than three units, the message had to repeat axes. Thus, an experimental message of four units contained the instructions to turn, climb, move, and turn again. In that respect, the experimental instructions differed from ATC instructions, in which each unit refers to a different action (e.g., an ATC message containing four elements could include instructions to change heading, altitude, speed, and radio frequency).

Figure 1. Computer display, sample instructions, and movements following instructions on the computer screen. (Note that during the experiment, the participants did not see either the sample instructions or the numbers indicating movements.)

The messages varied in length from one to six instructional units, or commands. For example, a message with one unit was "turn left one square," and a message with six units was "turn left one square, climb up two levels, move back one step, turn right three squares, climb down three levels, move forward two steps." Because of the consistency in terminology mentioned earlier (e.g., the word *turn* was always associated with *right* or *left* and with a number of *squares*), the words *turn*, *climb*, *move*, as well as the words *squares*, *levels*, and *steps*, were redundant in these messages. To test for the possible effects of such redundancy or wordiness, half of the messages were in the redundant four-words per unit (as in the two examples just given), and half were in a minimal two-words per unit format (i.e., "left one, up two, back one, right three, down three, forward two" for the six-unit example just given). There was an equal number of messages of each combination of length and wordiness.

Specifically, there were 72 experimental trials divided into six 12-trial blocks. The trials were presented in a pseudorandom order, with the constraint that each block of 12 trials included 1 trial of each length and wordiness combination. The six lengths were di-

vided into three length categories: short (one or two commands), medium (three or four commands), and long (five or six commands). Across the six blocks, for a given unit, each command (e.g., "turn left one square") was used equally often in each combination of wordiness and length category.

The experimental trials were preceded by specific instructions concerning how to follow the navigation commands, including a demonstration by the experimenter, using the small 3-D model. In addition, the computer provided a 1-trial demonstration, followed by 12 practice trials with feedback, including 1 of each combination of wordiness and length. The starting point for the practice trials was the mirror image of that for the experimental trials. The order of the practice trials was chosen in such a way that predicted difficulty increased systematically, unlike the experimental trials, in which predicted difficulty varied randomly.

Procedure

The participants heard messages of different lengths and different number of words per instructional unit. To turn (right or left) and to move (forward and back), the participants always stayed on

the same matrix. To climb (up or down), they went to a different matrix (but see Experiment 3). To follow the instructions, they clicked the computer mouse in the appropriate places in the grid. At the end of the instructions, the participants heard a beep. Following that beep sound, they repeated aloud the instructions they had heard. When they were done repeating, they clicked a button labeled DONE (see Figure 1). After clicking the button, they followed the instructions they had heard by clicking in the appropriate places on the grid. The participants had to click every square they passed, so if the instructions were "turn left two squares, climb down one level, move forward one step," they had to click on the square immediately to the left of the starting position (1 on Figure 1) and then on the square immediately to the left of that (2 on Figure 1), then on the square in the same location on the matrix below (3 on Figure 1), and once more on the square that was forward of their last position (4 on Figure 1). The number of clicks had to equal the number in the instructions (i.e., "two squares" = two clicks, "three steps" = three clicks, etc.). There was a 2-sec pause between trials.

EXPERIMENT 1

In Experiment 1, in a within-subjects design, we compared six different message lengths (one to six commands per message) and two different unit lengths, or wordiness levels (minimal and redundant), manipulated by varying the number of words per unit. We also compared the manual movement responses with the oral repetition responses. The wordiness manipulation was designed to test the hypothesis that the basic comprehension unit involved is propositional in nature, rather than word based. If propositional-like units are used rather than words, there should be a minimal effect of the wordiness manipulation. On the other hand, if words are the relevant units, participants should perform better with the minimal messages than with the redundant messages, because of the reduction in memory load. Such an effect of wordiness would seem more likely to occur with the oral repetition responses than with the manual movement responses, because the oral repetition responses resemble immediate recall of a list of words, where there are well-established list length effects (e.g., Murdock, 1962), as well as effects of output interference (e.g., Broadbent, 1958).

Method

Participants. Eight undergraduate students at the University of Colorado, Boulder, participated for credit in a course in introductory psychology. All the participants were native speakers of English.

Apparatus, Materials, and Procedure. The apparatus, procedure, and materials were the same as those described in the General Method section.

Analyses. Three multifactorial repeated measures analyses of variance (ANOVAs) were conducted for errors on the manual movement responses, as well as on the oral repetition responses. Each ANOVA included the factors of wordiness (minimal or redundant) and length (1–6 units). The first two involved a strict scoring procedure whereby a given trial was scored as correct only if the participant correctly followed all of the instructional units in the message; no partial credit was given. That is, an error was scored whenever the participant clicked in a wrong place on the grid or missed a click for the manual movement responses and whenever the participant said the wrong word or missed a word (out of the two critical words per instructional unit, as in the minimal messages) for the oral repetition responses. The responses, both manual and oral, had to follow the sequence given in the instructions to be scored as correct. The first

analysis included all six message lengths, whereas the second included only the intermediate lengths of three, four, and five units to avoid floor effects evident for lengths 1 and 2 and ceiling effects evident for length 6 (ceiling and floor in terms of errors). The third analysis involved a more lenient scoring procedure in which only errors on the first unit were examined; all six message lengths were included in this analysis. Because the pattern of the manual movement responses provided the foundation for understanding the importance of the pattern of the oral repetition responses, we report here (and in Experiments 2 and 3) the results of the manual movement responses before those of the oral repetition responses, despite the fact that the participants always gave the oral repetition responses before executing the manual movement responses. In addition, an overall ANOVA was conducted to test for differences between the manual movement responses and the oral repetition responses. Specifically, this last analysis involved strict scoring and included the within-subjects factor of response type (manual vs. oral), in addition to the two within-subjects factors (wordiness and length) examined in the separate analyses. Only effects involving the factor of response type are discussed for this last analysis.

Results

The results for the manual movement and oral repetition responses are summarized in Table 1 in terms of proportions of errors for the complete message, under both the strict and the lenient scoring methods, as a function of wordiness and length.

Manual movement responses. Errors increased dramatically with message length, from length 1 (.03) and length 2 (.03) to length 3 (.22) to length 4 (.52) to length 5 (.78) to length 6 [.90; $F(5,35) = 47.45$, $MS_e = 0.29$, $p < .001$]. Note that the largest drop in performance occurred between messages of length 3 and messages of length 4. A post hoc analysis comparing successive lengths yielded a significant effect of length pair [length 2–length 1 = .00; length 3–length 2 = .19; length 4–length 3 = .30; length 5–length 4 = .26; length 6–length 5 = .12; $F(4,28) = 3.46$, $MS_e = 0.03$, $p < .021$]. In contrast, there was no reliable effect of message wordiness.

In the analysis restricted to messages with three, four, or five units, there remained a significant main effect of length [$F(2,14) = 24.16$, $MS_e = 0.30$, $p < .001$].

Table 1
Proportion of Errors on Manual Movement and Oral Repetition Responses in Experiment 1 Under Strict and Lenient Scoring Methods as a Function of Wordiness and Length

Wordiness	Length					
	1	2	3	4	5	6
Strict Scoring, Manual Movement						
Minimal	.042	.062	.146	.500	.729	.937
Redundant	.021	.000	.292	.542	.833	.854
Lenient Scoring, Manual Movement						
Minimal	.021	.001	.042	.146	.208	.083
Redundant	.001	.001	.083	.104	.167	.083
Strict Scoring, Oral Repetition						
Minimal	.000	.000	.104	.375	.771	.979
Redundant	.000	.042	.333	.687	.854	.979
Lenient Scoring, Oral Repetition						
Minimal	.000	.000	.042	.104	.104	.167
Redundant	.000	.021	.104	.167	.250	.229

To determine whether message length has a list length effect (Murdock, 1962) in terms of the number of instructional units, we examined the proportion of errors on the first unit alone at each message length. These errors increased, although not monotonically, with length, from length 1 (.01) and length 2 (.00) to length 3 (.06) to length 4 (.13) to length 5 (.19) to length 6 [.08; $F(5,35) = 6.85$, $MS_e = 0.07$, $p < .001$]. The lower error rate on length 6 could be due to the fact that counterbalancing was done for each length category, rather than for each length. Errors did increase monotonically with length category, from short (.01) to medium (.09) to long (.14). In contrast, there was no reliable effect of message wordiness.

Oral repetition responses. As with the manual movement responses, there was a reliable effect of length. Errors increased dramatically and monotonically with message length, from length 1 (.00) to length 2 (.02) to length 3 (.22) to length 4 (.53) to length 5 (.81) to length 6 [.98; $F(5,35) = 105.54$, $MS_e = 0.16$, $p < .001$]. Again, the largest drop in performance was between messages of length 3 and messages of length 4. A post hoc analysis comparing successive lengths yielded a significant effect of length pair [length 2–length 1 = .02; length 3–length 2 = .20; length 4–length 3 = .31; length 5–length 4 = .28; length 6–length 5 = .17; $F(4,28) = 3.57$, $MS_e = 0.03$, $p < .019$].

In contrast to the manual movement responses, there was a reliable effect of message wordiness: Individuals erred less while repeating the minimal messages (.37) than while repeating the redundant messages [.48; $F(1,7) = 17.92$, $MS_e = 0.10$, $p < .005$], perhaps owing to a reduction in output interference (Broadbent, 1958) for the minimal messages. There was also a significant interaction of length and wordiness [$F(5,35) = 3.57$, $MS_e = 0.11$, $p < .011$], owing to floor and ceiling effects at the extreme message lengths (see the following analysis).

In the analysis restricted to messages containing three, four, or five units, there remained significant main effects of length [$F(2,14) = 38.71$, $MS_e = 0.22$, $p < .001$] and of message wordiness [$F(1,7) = 13.70$, $MS_e = 0.23$, $p < .008$], but there was no longer a reliable interaction of length and wordiness.

By the lenient scoring method, errors on the first unit increased dramatically and monotonically with message length, from length 1 (.00) to length 2 (.01) to length 3 (.07) to length 4 (.14) to length 5 (.18) to length 6 [.20; $F(5,35) = 7.63$, $MS_e = 0.09$, $p < .001$]. No other effects were significant in this analysis.

Combined analysis. The only significant effect involving response type was an interaction of wordiness and response type [$F(1,7) = 11.06$, $MS_e = 0.05$, $p < .013$], owing to the fact that the advantage of the minimal messages was greater for the oral repetition responses than for the manual movement responses and because the oral repetition responses showed an advantage in proportion of errors (.37) over the manual movement responses (.40) for the minimal messages, whereas the manual movement responses (.42) showed an advantage over the oral repetition responses (.48) for the redundant messages.

Discussion

Message length had the clearest effect on performance in this experiment. Not only did errors increase dramatically with message length, but also message length had a list length effect such that the longer the message, the more likely the participants were to make an error on the very first instructional unit. Thus, the effect of message length was not due only to the increase in performance requirement with each additional unit. Note that the largest drop in performance occurred between messages of length 3 and messages of length 4. This pattern was initially shown in the experiments by Barshi (1997), as well as in follow-up experiments by Barshi and Healy (1998). We return to this observation in the introduction to the next experiment and in the General Discussion section.

With respect to the number of words per unit, for the manual movement responses, the participants showed no cost from instructions that included more words per instructional unit (i.e., a greater redundancy). Importantly, doubling the number of words per unit had no reliable effect on the manual movement performance, whereas doubling the number of units had a large deleterious effect. This effect of increasing the number of units cannot be attributed simply to the increase in memory load and performance requirements, because we also found a large effect of message length when we examined errors on the first unit only. That is, for the first unit of a message, the number of subsequent units to be recalled had a significant effect on recall, whereas the number of words in the units themselves had no effect. Hence, the number of units per message is the primary determinant of comprehension, rather than the number of words per unit.

With respect to immediate recall performance as evidenced by the oral repetition responses, the minimal messages did have an advantage over the redundant messages, perhaps owing to a reduction in output interference on the minimal messages (Broadbent, 1958; Hadley, Healy, & Murdock, 1992). In other words, having to say more words per unit might have decreased the participants' ability to recall subsequent units. Interestingly, the additional words did not hinder the participants' subsequent manual movement responses.

Experiment 1 served as a benchmark for the remaining experiments reported here. It also replicated the main experimental results reported by Barshi (1997; Barshi & Healy, 1998), who examined only manual movement responses. These results included the large effect of message length, with its list length effect, and the large drop in performance between messages of length 3 and messages of length 4. Furthermore, this experiment showed nearly parallel patterns of performance on the manual and the oral repetition responses (except for effects of wordiness).

EXPERIMENT 2

In Experiment 2, in addition to the within-subjects factors included in Experiment 1 (message length, wordiness, and response type), a between-subjects factor of di-

mensions (3-D vs. 2-D) was employed. In this mixed design, we compared responses to instructions requiring movements along three dimensions (as in Experiment 1) with responses to instructions requiring movements along only two of the three dimensions (right/left and up/down). The advantage of choosing the vertical dimension as the second dimension for the 2-D condition was that in so doing, both conditions required a 3-D mental representation, since movement occurred outside of the picture plane.

The manipulation of the dimensions factor was similar in some ways to the experiments reported by Kerr (1993), in which the participants had to navigate in either a 2-D or a 3-D matrix. However, there were also several important differences between the design of the present experiment and that used by Kerr. In Kerr's experiments, movement was imagined with respect to a memory representation of the matrix, whereas in the experiments reported here, movement was physically performed on a visual display. Also, in Kerr's experiments, only arrival at the goal location was measured, whereas here, accuracy on both path and goal were measured. Most important, in Kerr's experiments, movement on the 2-D matrix was restricted to the picture plane, whereas the 2-D condition of the present experiment required movement on a plane orthogonal to the picture plane. Thus, the present experiment allowed us to determine whether the number of spatial dimensions through which navigation occurs influences navigation performance when a constant 3-D mental representation is employed.

Another theoretical issue addressed in this experiment concerns the claim by Barshi (1997), confirmed in the present Experiment 1, that the performance drop between three and four instructional units represents a cognitive limit reflecting short-term memory (STM) capacity. Because the three dimensions of the space allow for only three different commands (turn, climb, and move) and because of the fixed structure of the instructions, we had to start a second cycle through the same three dimensions to create messages of more than three instructional units in Experiment 1 (and in the earlier experiments by Barshi, 1997, and Barshi & Healy, 1998). Thus, it was possible that the large drop in performance we saw between messages of lengths 3 and 4 was an artifact of the experimental design, rather than a cognitive limit. In the 2-D condition of Experiment 2, we used only two dimensions along which movement could occur and so cycled through the same two dimensions every two instructional units, rather than every three units, as in Experiment 1 (and in the 3-D condition of Experiment 2); cycling through the dimensions is exemplified in the sample instructions and movements shown in Figure 2. If we were to find a large drop in performance between cycles (i.e., between messages of lengths 2 and 3 or lengths 4 and 5 in the 2-D condition), it would mean that the performance drop observed is an artifact of the experimental design. If, however, we were to find the same performance drop between messages of lengths 3 and 4

in both conditions of Experiment 2 (i.e., in mid-cycle in the 2-D condition), it would mean that the drop is indeed an indication of a cognitive limit.

Method

Participants. Twenty-four undergraduate students at the University of Colorado, Boulder, participated for credit in a course in introductory psychology. All the participants were native speakers of English.

Apparatus and Procedure. The apparatus and procedure were the same as those described in the General Method section.

Materials. The materials were the same as those described in the General Method section, except that the stimuli in the 2-D condition included only instructions to turn and to climb. For example, a message of six instructional units in the 2-D condition was "turn right two squares, climb up one level, turn left three squares, climb up one level, turn right three squares, climb down three levels" (see Figure 2). The stimulus sets for both conditions were identical in all other respects. Also, these stimulus sets were slightly different from the one used in Experiment 1; the new sets were made to meet tighter constraints such that counterbalancing was obtained across all message lengths, and not just across length categories, as in Experiment 1. Specifically, across the 72 trials in a given condition, for a given unit, each command was used equally often in each combination of wordiness and length.

Analyses. A multifactorial repeated measures ANOVA was conducted for the manual movement responses, as well as for the oral repetition responses. Each ANOVA included the between-subjects factor of dimension (3-D or 2-D) and the within-subjects factors of wordiness and length. In addition, an overall ANOVA was conducted to test for differences between the manual movement responses and the oral repetition responses. This last analysis included the additional within-subjects factor of response type. All the analyses involved the strict scoring procedure and included all six message lengths. Given that the list length effect was established clearly in Experiment 1, for the sake of brevity, the analysis involving lenient scoring is not reported either here or in Experiment 3.

Results

The results for the manual movement and oral repetition responses are summarized in Table 2 in terms of proportions of errors on the complete message, under the strict scoring method, as a function of dimension, wordiness, and length.

Manual movement responses. Errors increased dramatically and monotonically with message length, from length 1 (.02) to length 2 (.12) to length 3 (.33) to length 4 (.63) to length 5 (.81) to length 6 [.91; $F(5,110) = 151.39$, $MS_e = 0.26$, $p < .001$]. This pattern shows again that the largest drop in performance occurred between messages of length 3 and messages of length 4. A post hoc analysis comparing successive lengths yielded a significant effect of length pair [length 2–length 1 = .09; length 3–length 2 = .21; length 4–length 3 = .30; length 5–length 4 = .18; length 6–length 5 = .10; $F(4,22) = 4.11$, $MS_e = 0.04$, $p < .005$].

In contrast, there was no main effect of wordiness. However, wordiness did interact reliably with length [$F(5,110) = 2.43$, $MS_e = 0.12$, $p < .039$], owing to the fact that minimal messages had an overall advantage for lengths 3 and 4, whereas redundant messages had an ad-

Figure 2. Sample instructions and movements in the three-dimensional condition (top panel) and the two-dimensional condition (bottom panel) of Experiment 2. (Note that during the experiment, the participants did not see either the sample instructions or the numbers indicating movements.)

Table 2
Proportion of Errors on Manual Movement and Oral Repetition Responses
in Experiment 2 as a Function of Dimension, Wordiness, and Length

Dimension	Wordiness	Length					
		1	2	3	4	5	6
Manual Movement							
Three-dimensional	minimal	.014	.153	.292	.653	.889	.958
	redundant	.028	.069	.403	.722	.847	.931
Two-dimensional	minimal	.028	.139	.264	.556	.806	.875
	redundant	.014	.097	.347	.583	.708	.875
Oral Repetition							
Three-dimensional	minimal	.000	.000	.181	.500	.792	1.000
	redundant	.000	.042	.278	.722	.819	.944
Two-dimensional	minimal	.000	.000	.153	.472	.764	.875
	redundant	.000	.139	.333	.597	.736	.847

vantage for all other message lengths. Importantly, there was no main effect of the between-subjects factor of dimension [$F(1,22) < 1$] and no reliable interactions involving the between-subjects factor.

Oral repetition responses. As with the manual movement responses, there was no main effect of the between-subjects factor of dimension [$F(1,22) < 1$] and no reliable interactions involving that factor. Errors increased dramatically and monotonically with message length, from length 1 (.00) to length 2 (.05) to length 3 (.24) to length 4 (.57) to length 5 (.78) to length 6 [.92; $F(5,110) = 187.62$, $MS_e = 0.22$, $p < .001$]. Here, too, the largest drop in performance was between messages of length 3 and length 4. A post hoc analysis comparing successive lengths yielded a significant effect of length pair [length 2–length 1 = .05; length 3–length 2 = .19; length 4–length 3 = .34; length 5–length 4 = .21; length 6–length 5 = .14; $F(4,22) = 7.71$, $MS_e = 0.04$, $p < .001$].

In contrast to the manual movement responses, but as in Experiment 1, there was a reliable effect of message wordiness: Individuals erred less while repeating the minimal messages (.40) than while repeating the redundant messages [.46; $F(1,22) = 11.77$, $MS_e = 0.13$, $p < .003$], perhaps owing to a reduction in output interference for the minimal messages. There was a significant interaction of length and wordiness [$F(5,110) = 4.54$, $MS_e = 0.12$, $p < .002$], owing to the fact that the advantage of the minimal messages was greatest for messages of lengths 3 and 4 and because of floor and ceiling effects at the extreme message lengths.

Combined analysis. In contrast to Experiment 1, there was a reliable main effect of response type: Individuals erred less while orally repeating the messages (.43) than while following the instructions manually [.47; $F(1,22) = 10.36$, $MS_e = 0.16$, $p < .005$]. Wordiness and response type also interacted reliably [as in Experiment 1; $F(1,22) = 14.83$, $MS_e = 0.05$, $p < .002$], owing to the fact that minimal messages had an advantage in proportion of errors over redundant messages only for the oral repetition responses (minimal = .40; redundant = .46), not for the manual responses (minimal = .47; redundant = .47). In addition, there was a reliable three-way interaction involving

wordiness, length, and response type [$F(5,110) = 2.88$, $MS_e = 0.06$, $p < .018$]. However, no clear pattern is discernible in this higher order interaction. As with the separate analyses, there was no main effect of the between-subjects factor of dimension [$F(1,22) < 1$] and no reliable interactions involving the between-subjects factor.

Discussion

Experiment 2 replicated our previous results with respect to the effect of message length and the advantage of minimal messages for the oral repetition responses only. The fact that we replicated in this experiment the previously found large drop in performance (relative to the rest of the function) between messages of lengths 3 and 4 even in the 2-D condition shows that this sharp performance decline is due to cognitive limitations and is not an artifact of the experimental design.

Interestingly, we did not find an overall effect of number of dimensions. That is, responding to instructions (both orally and manually) requiring movement along only two dimensions of a 3-D space was not easier than responding to instructions requiring movement along all three dimensions. Given that both dimension conditions (3-D and 2-D) required movement outside the picture plane, this experiment suggests that it is the mental representation that is critical to performance, rather than the number of dimensions along which movement is required. Experiment 3 tested this suggestion further by varying the spatial representation required. In addition, Experiment 3 explored the relationship between verbal and spatial memory processes by examining whether the spatial representation affects immediate memory for the verbal navigation instructions.

EXPERIMENT 3

In Experiment 3, in addition to the within-subjects factors included in Experiments 1 and 2 (message length, wordiness, and response type), a between-subjects factor of matrix (single vs. multi) was employed. In this mixed design, the instructions for both matrix conditions required movements along two dimensions only (as in the

2-D condition of Experiment 2). However, in one condition (single-matrix), movements were to be performed within a single matrix only and thus, presumably, required only a 2-D mental representation, because all movements were performed on the picture plane. In the other condition (multi-matrix), movements involved all four matrices of the grid and thus, presumably, required a 3-D mental representation, because movements were performed on a plane orthogonal to the picture plane. The multi-matrix condition of Experiment 3 was a replication of the 2-D condition of Experiment 2. This experiment allowed us to determine whether the additional complexity of a 3-D, as opposed to a 2-D, spatial representation impedes performance on the navigation task.

Because the verbal navigation instructions were identical in the two matrix conditions, differences between the single-matrix and the multi-matrix conditions should be restricted to the manual movement responses if verbal and spatial memory processes are independent of each other, as has been suggested by some studies of WM (Friedman & Miyake, 2000; Shah & Miyake, 1996; Toms et al., 1994). On the other hand, if verbal and spatial memory processes are interdependent, as was sug-

gested by the work of McNamara and LeSueur (1989), differences between the single- and the multi-matrix conditions should be observed also in the oral repetition responses that preceded manual movement.

Method

Participants. Twenty undergraduate students at the University of Colorado, Boulder, participated for credit in a course in introductory psychology. All the participants were native speakers of English.

Apparatus. The apparatus was the same as that described in the General Method section.

Materials. The materials were the same as those described for the 2-D condition of Experiment 2; the words used in the instructions were the same in the multi-matrix and the single-matrix conditions.

Procedure. The procedure was the same as that described in the General Method section, except that in the single-matrix condition, movements were to take place within a single matrix. For example, in response to an instruction to "climb down one level," the participants were to click in the square immediately below the square they were on within the same matrix (see the left panel of Figure 3). In response to the same instruction, the participants in the multi-matrix condition were to click on the square of the same location as the one they were on in the matrix below the one they were on (see the right panel of Figure 3).

Analyses. A multifactorial repeated measures ANOVA was conducted for the manual movement responses, as well as for the oral

Figure 3. Sample instructions and movements in the single-matrix condition (left panel) and the multi-matrix condition (right panel) of Experiment 3. (Note that during the experiment, the participants did not see either the sample instructions or the numbers indicating movements.)

repetition responses. Each ANOVA included the between-subjects factor of matrix condition (single, multi) and the within-subjects factors of wordiness and length. In addition, an overall ANOVA was conducted to test for differences between the manual movement responses and the oral repetition responses. This last analysis included the additional within-subjects factor of response type. As with Experiment 2, we report here only the analyses involving the strict scoring procedure.

Results

The results for the manual movement and oral repetition responses are summarized in Table 3 in terms of proportions of errors on the complete message, under the strict scoring procedure, as a function of matrix condition, wordiness, and length.

Manual movement responses. There was a reliable main effect of matrix condition [$F(1,18) = 12.38$, $MS_e = 0.62$, $p < .003$]. The participants made more errors overall in the multi-matrix condition (.44) than in the single-matrix condition (.30). The between-subjects factor of matrix condition also interacted reliably with the within-subjects factor of wordiness [$F(1,18) = 4.78$, $MS_e = 0.09$, $p < .041$], owing to the fact that minimal messages had no advantage (.44) over redundant messages (.44) in the multi-matrix condition, whereas redundant messages had an advantage (.26) over minimal messages (.33) in the single-matrix condition. There was also a reliable interaction involving matrix condition and length [$F(5,90) = 4.73$, $MS_e = 0.16$, $p < .001$], owing to the fact that the difference between the two matrix conditions increased as message length increased (see the rows labeled "M" in Table 3).

There was also a reliable main effect of length. Errors increased dramatically and monotonically with message length, from length 1 (.00) to length 2 (.04) to length 3 (.20) to length 4 (.45) to length 5 (.70) to length 6 [.82; $F(5,90) = 171.97$, $MS_e = 0.16$, $p < .001$]. As previously, the largest numerical drop in performance was between messages of length 3 and length 4, although it was equivalent to that be-

tween messages of length 4 and length 5 by rounding. A post hoc analysis comparing successive lengths yielded a significant effect of length pair [length 2–length 1 = .03; length 3–length 2 = .16; length 4–length 3 = .25; length 5–length 4 = .25; length 6–length 5 = .12; $F(4,18) = 7.41$, $MS_e = 0.02$, $p < .001$]. There was no reliable main effect of wordiness.

Oral repetition responses. Importantly, as with the manual movement responses, there was a reliable main effect of matrix condition [$F(1,18) = 7.31$, $MS_e = 0.82$, $p < .014$]. The participants made more errors overall in the multi-matrix condition (.41) than in the single-matrix condition (.28), even though the words spoken were identical in the two matrix conditions. The between-subjects factor of matrix also interacted reliably with the within-subjects factor of length [$F(5,90) = 6.57$, $MS_e = 0.17$, $p < .001$], owing to the fact that the difference between the two matrix conditions increased as message length increased.

Errors increased dramatically and monotonically with message length, from length 1 (.00) to length 2 (.03) to length 3 (.19) to length 4 (.40) to length 5 (.66) to length 6 [.81; $F(5,90) = 162.54$, $MS_e = 0.17$, $p < .001$]. In this case, in which overall performance was relatively high, the largest drop in performance occurred between messages of lengths 4 and 5. A post hoc analysis comparing successive lengths yielded a significant effect of length pair [length 2–length 1 = .03; length 3–length 2 = .17; length 4–length 3 = .21; length 5–length 4 = .26; length 6–length 5 = .15; $F(4,18) = 5.60$, $MS_e = 0.03$, $p < .001$].

There was also a significant interaction of length and wordiness [$F(5,90) = 2.74$, $MS_e = 0.11$, $p < .024$], owing to the fact that redundant messages had an advantage for messages of lengths 5 and 6, whereas minimal messages had an advantage for messages of lengths 2, 3, and 4.

Combined analysis. In contrast to the results of Experiment 2, but consistent with the results of Experi-

Table 3
Proportion of Errors on Manual Movement and Oral Repetition Responses
in Experiment 3 as a Function of Matrix Condition, Wordiness, and Length

Matrix Condition	Wordiness	Length					
		1	2	3	4	5	6
Manual Movement							
Multi	minimal	.000	.083	.233	.533	.833	.950
	redundant	.000	.033	.283	.583	.817	.950
	<i>M</i>	.000	.058	.258	.558	.825	.950
Single	minimal	.000	.017	.200	.383	.600	.767
	redundant	.017	.017	.083	.317	.550	.600
	<i>M</i>	.008	.017	.142	.350	.575	.683
Oral Repetition							
Multi	minimal	.000	.000	.150	.433	.800	.983
	redundant	.000	.033	.300	.533	.767	.950
	<i>M</i>	.000	.017	.225	.483	.783	.967
Single	minimal	.000	.000	.100	.317	.550	.700
	redundant	.000	.067	.217	.317	.533	.600
	<i>M</i>	.000	.033	.158	.317	.542	.650

ment 1, there was no overall difference in performance between the manual movement responses and the oral repetition response (i.e., no main effect of response type). There was a reliable interaction of wordiness and response type [as in Experiments 1 and 2; $F(1,18) = 6.80$, $MS_e = 0.07$, $p < .017$], owing to the fact that minimal messages (.34) had an advantage in proportion of errors over redundant messages (.36) for the oral repetition responses, whereas redundant messages (.35) had an advantage over minimal messages (.38) for the manual movement responses. In addition, there was a significant, but complex, three-way interaction of wordiness, length, and response type [$F(5,90) = 3.17$, $MS_e = 0.04$, $p < .012$].

Discussion

Experiment 3 replicated our previous finding of the effect of message length. It also showed that whatever advantage minimal messages might have, it is limited to the oral repetition responses (and to short message lengths).

Experiment 3 contrasts with Experiment 2 in a very crucial way. Recall that in Experiment 2, we found no difference between the 3-D and the 2-D conditions. Recall also that the multi-matrix condition of Experiment 3 was identical to the 2-D condition of Experiment 2 and that the participants heard exactly the same words in both matrix conditions of Experiment 3. Nevertheless, we found a reliable difference between the two matrix conditions in Experiment 3 (even though there were fewer participants in Experiment 3 than in Experiment 2): The participants in the single-matrix condition made fewer errors than did the participants in the multi-matrix condition, and this difference was clear for the oral repetition responses, as well as for the manual movement responses.

We attributed the lack of difference between the two conditions of Experiment 2 to the fact that both of them required a 3-D mental representation because movement occurred outside the picture plane in both conditions. In the present experiment, however, only the multi-matrix condition required movement outside the picture plane; all the movement in the single-matrix condition occurred within the picture plane. Thus, it seems that the multi-matrix condition required a 3-D mental representation, whereas the single-matrix condition required only a 2-D mental representation, even though the words to which the participants responded were identical in both conditions.

Finding an effect of matrix condition on the oral repetition responses is consistent with the view that verbal and spatial memory processes are interdependent. Thus, the reliable difference between the two matrix conditions in the present experiment suggests that the mental representation of the space to which the language input refers plays a crucial role in comprehension. The fact that the participants' ability simply to repeat the words they heard was significantly different in the two matrix conditions suggests that this spatial representation also plays a role in immediate verbatim recall.

GENERAL DISCUSSION

The present research has probed a number of aspects of the mental representations necessary to perform a navigation task and the effects of those representations on performance in that task. Experiment 1 suggested that comprehension is affected by the number of instructional units, rather than by the number of words per unit. The results of Experiment 2 showed that when moving in a 3-D space, it does not matter whether movement is required along all three dimensions or along only two of the three dimensions. The results of Experiment 3 showed that individuals performed much better when they had to maintain a 2-D mental representation than when they had to maintain a 3-D mental representation. What is more, it showed that even immediate verbatim recall was affected by the spatial representation of the situation to which the language input applied.

We found that wordiness had no overall effect on the manual movement responses. However, we found that in Experiments 1 and 2, wordiness had a main effect on the oral repetition responses such that minimal messages led to more accurate repetitions than did redundant messages. The oral repetition responses were a verbatim recall task and were subject to output interference (Broadbent, 1958; Hadley et al., 1992) such that the more words the participants had to repeat, the more likely they were to make errors independent of the propositional content.¹ It is the accuracy of the manual movement responses that provided a strict measure of comprehension, and it is there that we find a striking pattern.

Although doubling the number of instructional units severely reduced movement performance, doubling the number of words in each unit generally had no effect on movement accuracy. For example, the minimal (i.e., with two words per unit) messages containing two units had a total of four words, whereas the redundant (i.e., with four words per unit) messages containing two units had a total of eight words. This manipulation of going from four to eight words per message had no effect on movement responses. In contrast, the minimal message containing two units had a total of four words in it, whereas the minimal message containing four units had a total of eight words in it. This manipulation of going from four to eight words per message yielded a dramatic decrease in accuracy of the manual movement responses. Thus, it is the number of units, not the number of words, that is the primary determinant of comprehension. This finding cannot be attributed simply to the fact that, with more units, there are more movements to be made, because there was a large effect of the number of units even when errors only on the first unit were examined. Finding a list length effect in this case is notable because examining errors on the first unit only is equivalent to performing an analysis of accuracy for only the first serial position of a list, and list length effects have been shown to be minimal in such cases (Drewnowski & Murdock, 1980).

Instructional units, as used in the experimental stimuli, can be considered to be single propositions. Operationally, they were clearly single chunks, even though they contained two discrete pieces of information: direction and magnitude. Movement along each axis had to include both elements, and so each instructional unit corresponded to a single axis. In establishing the role of the unit, over and above the role of the individual words, in the memory for and the comprehension of the navigation instructions, the results of Experiment 1 go beyond the classic findings of Sachs (1967), who showed that individuals demonstrate long-term memory for meaning but lose surface features of utterances rapidly, and beyond the findings of Kintsch and Keenan (1973), who argued for the propositional nature of semantic representation. Kintsch and Keenan compared reading rates and retention of sentences that varied in the number of words and the number of propositions and found that "the number of propositions, although by no means the whole story, has proven to be a useful independent variable for the analysis of both reading time and recall data" (p. 272). However, they did not compare the same sentences, with the same propositional content, containing the same number of propositions but varying only the number of words, as was done in the present experiment. Furthermore, Kintsch and Keenan clearly stated that many important variables, such as syntactic and semantic complexity, number and nature of arguments, and structural relations were not controlled or manipulated in their study. In contrast, the simple design of the materials in Experiment 1 allowed us to control for all these factors, hold the propositional content of each unit fixed, and vary only the number of words. Thus, the results of the wordiness manipulation reported here strongly suggest that the unit in question is propositional in nature.

We found that length in terms of the number of units had a large and consistent effect on both the oral repetition and the manual movement responses. The effect of length was particularly evident in the comparison of instructions with three and four units, even in the 2-D conditions in Experiments 2 and 3, which ruled out the possibility that this effect was an artifact of the experimental design. The drop in performance between successive lengths was largest for lengths 3 and 4 in each experiment, with the exception of the oral repetition responses in Experiment 3, in which overall performance was relatively high and the largest drop in performance was, instead, between messages of lengths 4 and 5. The otherwise consistent trend is best exemplified for the oral repetition responses in Experiment 2, in which Newman-Keuls paired comparisons yielded a significant difference between the drop from length 3 to length 4 and the drops between all other successive pairs.

These results are consistent with those of Barshi (1997; see also Barshi & Healy, 1998), who argued that the drop in performance between messages of lengths 3 and 4 pointed to a WM capacity that was limited to three chunks under a more naturalistic setting than the pure

STM laboratory task because it involved processing as well as storage requirements. Barshi (1997) also observed an interesting effect of practice on the differences in performance between messages of lengths 3 and 4: The learning that was evident with length 3 did not occur with length 4. Often, with the same amount of practice, the participants reached error levels of close to 0% on length 3, but only about 50% on length 4. This pattern of performance provides evidence for setting the upper limit of WM capacity at three units.

Further support for this capacity limit comes from a different type of data, also involving navigation instructions. Barshi (1997) examined misunderstandings in communication between pilots and air traffic controllers. His analysis of actual recordings of the natural language used in ATC communication revealed that the deletion of linguistic elements, such as prepositions, determiners, and sequence markers, had no reliable effect on the comprehension of ATC instructions. In addition, he showed, both in the natural ATC data and in the results of four experiments, that prosody and speech rate had no direct effect on comprehension, whereas message length, in terms of the number of aviation topics (Prinzo, Britton, & Hendrix, 1996), had a large effect. There was a large increase in the number of misunderstandings when ATC instructions included more than three aviation topics. These findings were consistent with findings reported by Morrow and Rodvold (1993), who found, in a simulation study, that misunderstandings were more likely when a message contained four commands than when the same four commands were given in two separate messages, each containing two commands only, and each message followed by a separate response. Barshi's (1997) findings were also consistent with those of Cardosi and her collaborators (e.g., Cardosi, 1993, 1994; Cardosi, Brett, & Han, 1996), who studied natural ATC communication and observed that "while clearances that contained four or more pieces of information made up only 26% of the readbacks, they accounted for 51% of the readback errors found in this study" (Cardosi et al., 1996, p. 9).

From both manipulations of wordiness and length, we conclude that the capacity limit of three units identified by Barshi (1997) corresponds to a limit of three propositions. Placing the language comprehension capacity limit of WM around three is also supported by a range of diverse studies (see, e.g., Broadbent, 1975; Cowan, 2001), including one in which the average span for unpronounceable Chinese radicals was set at 3.0 (Zhang & Simon, 1985) and one in which the average reading span was set at 3.15 (Daneman & Carpenter, 1980). Such studies led Lewis (1996) to argue that for sentence processing, the "magical number" is, in fact, two or three. Is this number also "magical" with respect to the number of dimensions?

According to Kerr (1993), array theories of imagery (e.g., Kosslyn, 1980) argue that visual perception contains only two dimensions and that, to visualize the third

dimension, individuals must extract information about that third dimension from the other two dimensions available from the visual perception process. Therefore, array theories predict that handling a 3-D space ought to be more difficult than handling a 2-D one. Experiments 2 and 3 show that the part that is difficult in handling a 3-D space is its mental representation. This conclusion follows from the fact that Experiment 2 showed no difference in performance between the 3-D condition and the 2-D condition, whereas Experiment 3 showed a reliable performance advantage for the single-matrix condition over the multi-matrix condition with a smaller sample size ($n = 20$ in Experiment 3 and $n = 24$ in Experiment 2). In the 2-D condition of Experiment 2, movement was limited to only two of the three dimensions of the space, but the space had to be seen as a 3-D one because movement occurred outside the picture plane. The same was true for the multi-matrix condition of Experiment 3, which was an exact replication of the 2-D condition of Experiment 2. However, in the single-matrix condition of Experiment 3, all movement occurred within the picture plane. Thus, even though the words in the navigation instructions in the two matrix conditions of Experiment 3 were identical, the single-matrix condition allowed for a 2-D mental representation, whereas the multi-matrix condition (and the 2-D condition of Experiment 2) required a 3-D mental representation, as was the case in the 3-D condition of Experiment 2. The performance advantage seen for the single-matrix condition would seem, therefore, consistent with an array theory, rather than a sandbox theory, of imagery (e.g., Attneave, 1972), because sandbox theories (according to Kerr, 1993) postulate that all three dimensions are perceived directly and so, no particular difficulty should arise in handling a 3-D space, as compared with handling a 2-D space.

Kerr (1993) interprets array theories as considering depth as the critical dimension. She says that "array theories would predict slower imagery processing time whenever the depth dimension was included in a matrix, even if the matrix were itself essentially two-dimensional (e.g., a 3×3 matrix that was only one unit wide, but three units high and three units deep)" (p. 472). Depth was also addressed by Bryant and Tversky (1999), who found that adding depth cues to 2-D diagrams caused the participants to switch from an intrinsic computation analysis, whereby they viewed the scene from the outside (without depth cues), to a spatial framework analysis, whereby they viewed the scene from the perspective of the character within it (with depth cues). The movements required in the multi-matrix condition of Experiment 3 did not involve the depth dimension. Instead, in that condition, movements were made only along the dimensions of height and width. The results of the present study, then, would not support any special status for the depth dimension, because performance was consistently worse in the multi-matrix condition than in the single-matrix condition. Rather, the present results imply that what makes one spatial representation more complex

and harder to maintain and manipulate than another spatial representation is the existence of a dimension that is outside the picture plane (regardless of which dimension it is). This implication should also be relevant to other studies involving imagery/perceptual representations in which three versus two dimensions are compared. For example, Bauer and Jolicœur (1996) studied mental rotation and compared both 3-D versus 2-D rotations and rotations inside versus outside the picture plane. They found that "3-D figures result in slower rotation rates than do 2-D figures" (p. 92). However, they also found that rotations outside the picture plane (along the depth axis) were faster than those within the picture plane. The latter finding seems inconsistent both with Kerr's conclusions concerning the difficulty of processing the depth dimension and our conclusions concerning the difficulty of processing outside the picture plane. The inconsistency between the conclusions of Bauer and Jolicœur and the present conclusions suggests that both depth and the picture plane may play different roles in imagery/perceptual representations than in the conceptual representations examined in the present study (but see Bryant, 1997, for a discussion of a single cognitive system for constructing spatial representations from linguistic or perceptual inputs).

Regardless of the nature of the spatial representation adopted and maintained, the results of the present study suggest that the spatial representation definitely affects performance, even with a task that has traditionally been considered a measure of verbal STM—immediate verbatim recall. Although immediate verbatim recall could be based on purely verbal memory, our finding that the spatial representation required by the manual movement responses affected immediate verbatim recall implies that the recall in that case, in fact, did not depend solely on verbal memory. The performance advantage we found for the single-matrix condition over the multi-matrix condition in the oral repetition responses of Experiment 3 clearly shows that the nature of the mental representation of the space to which the instructions applied affected individuals' ability simply to repeat back the instructions they heard even though the words were identical in the two conditions. Thus, the spatial representation affected memory for the verbal material (i.e., the verbal representation). These findings provide strong support for the interdependence of the spatial and the verbal memory processes, as had been suggested by the work of McNamara and LeSueur (1989); the findings are not consistent with the notion that spatial and verbal WM processes are completely independent, as had been suggested by some previous investigations of WM (Friedman & Miyake, 2000; Shah & Miyake, 1996; Toms et al., 1994).

These findings are also relevant to discussions of situation/mental models (for a review, see Zwaan & Radvansky, 1998), in which it has been assumed that new input is integrated into an existing situation model or that it is used in the construction of such a model. However,

most studies that have employed memory tests to investigate situation models have focused on delayed memory tasks (as in the case of text comprehension; e.g., Kintsch, 1988) and have not shown that a particular situation model has any effects on immediate memory tasks. Because the words to be recalled in the two matrix conditions were identical, the large difference in immediate verbatim recall performance between the two conditions could be explained only by the differences between the situations (namely, their mental representations) to which the language input was relevant.

In summary, our study was able to shed light on some aspects of the verbal representation, the spatial representation, and the relation between the two representations in the task of following navigation instructions. We found evidence that the units in the verbal representation of this task are propositional in nature, rather than word based. We found that the number of spatial dimensions required in the spatial representation of this task had a strong effect on performance, whereas the number of traveled dimensions had no effect. We also found that the spatial representation affected immediate verbatim recall, as well as movement performance (i.e., both memory and comprehension), strongly supporting the interdependence of the verbal and the spatial representations underlying this task.

On the basis of the research reported here, we can also offer some practical recommendations, reflecting the effects of each of the three major variables we examined: message length, message wordiness, and the dimensionality of the space. First, with respect to message length, this study supports the recommendation made by Barshi (1997) that to minimize errors resulting from message length, it would be advisable for air traffic controllers to limit their transmissions to no more than three aviation topics. This recommendation could be easily extended to any verbal instruction-giving context: To minimize errors, instructors should avoid giving more than three instructional units in any one utterance. Second, with respect to message wordiness, given that there was no cost of redundancy in executing the instructions, to minimize errors it would be advisable to maintain message redundancy, especially in cases in which ambiguity could arise (e.g., when the number 210 in ATC instructions could refer to altitude, heading, or airspeed). Third, with respect to spatial dimensionality, because of the difficulties individuals face when having to maintain a 3-D mental representation, additional training is advisable whenever a target task requires such a representation.

REFERENCES

- ATTNEAVE, F. (1972). Representation of physical space. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 283-306). Washington, DC: Winston.
- BARSHI, I. (1997). *Effects of linguistic properties and message length on misunderstandings in aviation communication*. Unpublished doctoral dissertation, University of Colorado, Boulder.
- BARSHI, I., & HEALY, A. F. (1998). Misunderstandings in voice communication: Effects of fluency in a second language. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 161-192). Hillsdale, NJ: Erlbaum.
- BAUER, B., & JOLICŒUR, P. (1996). Stimulus dimensionality effects in mental rotation. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 82-94.
- BJORK, E. L., & HEALY, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning & Verbal Behavior*, **13**, 80-97.
- BROADBENT, D. E. (1958). *Perception and communication*. New York: Pergamon.
- BROADBENT, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (pp. 3-18). London: Wiley.
- BRYANT, D. J. (1997). Representing space in language and perception. *Mind & Language*, **12**, 239-264.
- BRYANT, D. J., & TVERSKY, B. (1999). Mental representations of perspective and spatial relations from diagrams and models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 137-156.
- CARDOSI, K. M. (1993). *An analysis of en route controller-pilot voice communication* (NTIS No. PB93-189702/HDM). Cambridge, MA: John A. Volpe National Transportation Systems Center.
- CARDOSI, K. M. (1994). *An analysis of tower (local) controller-pilot voice communications* (NTIS No. PB94-217072/HDM). Cambridge, MA: John A. Volpe National Transportation Systems Center.
- CARDOSI, K. M., BRETT, B., & HAN, S. (1996). *An analysis of TRACON (terminal radar approach control) controller-pilot voice communications* (NTIS No. PB96-202593/HDM). Cambridge, MA: John A. Volpe National Transportation Systems Center.
- COWAN, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, **24**, 87-185.
- DANEMAN, M., & CARPENTER, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, **19**, 450-466.
- DREWNOWSKI, A., & MURDOCK, B. B., JR. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 319-332.
- ESTES, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161-190). New York: Halsted Press.
- FRANKLIN, N., & TVERSKY, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, **119**, 63-76.
- FRIEDMAN, N. P., & MIYAKE, A. (2000). Differential roles for visuo-spatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General*, **129**, 61-83.
- GARDEN, S., CORNOLDI, C., & LOGIE, R. H. (2002). Visuo-spatial working memory in navigation. *Applied Cognitive Psychology*, **16**, 35-50.
- HADLEY, J. A., HEALY, A. F., & MURDOCK, B. B. (1992). Output and retrieval interference in the missing-number task. *Memory & Cognition*, **20**, 69-82.
- HEALY, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning & Verbal Behavior*, **13**, 644-655.
- JOHNSON-LAIRD, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
- KERR, N. H. (1993). Rate of imagery processing in two versus three dimensions. *Memory & Cognition*, **21**, 467-476.
- KINTSCH, W. (1972). Notes on the structure of semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 247-308). New York: Academic Press.
- KINTSCH, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, **95**, 163-182.
- KINTSCH, W. (1994). The psychology of discourse processing. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 721-739). San Diego: Academic Press.
- KINTSCH, W., & KEENAN, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, **5**, 257-274.
- KOSSLYN, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- LEWIS, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, **25**, 93-115.
- MCMANARA, T. P., & LESUEUR, L. L. (1989). Mental representations of

- spatial and nonspatial relations. *Quarterly Journal of Experimental Psychology*, **41A**, 215-233.
- MORROW, D. G., BOWER, G. H., & GREENSPAN, S. L. (1989). Updating situation models during narrative comprehension. *Journal of Memory & Language*, **28**, 292-312.
- MORROW, D. G., GREENSPAN, S. L., & BOWER, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory & Language*, **26**, 165-187.
- MORROW, D. G., & RODVOLD, M. (1993). *The influence of ATC message length and timing on pilot communication* (NTIS No. N94-14744). Moffett Field, CA: NASA Ames Research Center.
- MURDOCK, B. B., JR. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, **64**, 482-488.
- PERRIG, W., & KINTSCH, W. (1985). Propositional and situational representations of text. *Journal of Memory & Language*, **24**, 503-518.
- PRINZO, O. V., BRITTON, T. W., & HENDRIX, A. M. (1996). Development of a coding form for approach control/pilot voice communication. In B. G. Kanki & O. V. Prinzo (Eds.), *Methods and metrics of voice communication* (pp. A-51-A-76). Washington, DC: DOT, FAA, Office of Aviation Medicine.
- RAPP, D. N., GERRIG, R. J., & PRENTICE, D. A. (2001). Readers' trait-based models of characters in narrative comprehension. *Journal of Memory & Language*, **45**, 737-750.
- SACHS, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, **2**, 437-442.
- SHAH, P., & MIYAKE, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, **125**, 4-27.
- TOMS, M., MORRIS, N., & FOLEY, P. (1994). Characteristics of visual interference with visuospatial working memory. *British Journal of Psychology*, **85**, 131-144.
- VAN DIJK, T. A., & KINTSCH, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- ZHANG, G., & SIMON, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and the acoustical loop hypotheses. *Memory & Cognition*, **13**, 193-201.
- ZWAAN, R. A., MADDEN, C. J., & WHITTEN, S. N. (2000). The presence of an event in the narrated situation affects its availability to the comprehender. *Memory & Cognition*, **28**, 1022-1028.
- ZWAAN, R. A., & RADVANSKY, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, **123**, 162-185.

NOTE

1. A set of post hoc serial position error analyses was conducted in order to examine more directly output interference effects in the oral repetition responses. In accordance with predictions based on output interference, for each of the three experiments, these analyses yielded rising serial position functions (i.e., increasing error rates or declining performance) for each message length. These functions were most dramatic for the longer messages. For example, there were large increases in error rate across serial positions with messages of length 6 in each experiment [Experiment 1, Position 1 = .20, Position 2 = .25, Position 3 = .69, Position 4 = .73, Position 5 = .77, Position 6 = .90, $F(5,35) = 30.68$, $MS_e = 0.27$, $p < .001$; Experiment 2, Position 1 = .21, Position 2 = .27, Position 3 = .53, Position 4 = .55, Position 5 = .76, Position 6 = .83, $F(5,110) = 67.65$, $MS_e = 0.26$, $p < .001$; Experiment 3, Position 1 = .21, Position 2 = .20, Position 3 = .36, Position 4 = .46, Position 5 = .59, Position 6 = .73, $F(5,90) = 91.60$, $MS_e = 0.12$, $p < .001$]. Although these serial position functions do not have the typical bow shape found in studies of immediate memory (e.g., Estes, 1972), bow-shaped functions have been found to be limited to order errors (e.g., Bjork & Healy, 1974; Healy, 1974), which were greatly reduced or eliminated in the present study by the fixed order of the units in the messages.

(Manuscript received February 27, 2001;
revision accepted for publication August 21, 2002.)