

Chunking in recall of symbolic drawings

DENNIS E. EGAN and BARRY J. SCHWARTZ
Bell Laboratories, Murray Hill, New Jersey 07974

Three experiments explored memory for symbolic circuit drawings using skilled electronics technicians and novice subjects. In the first experiment a skilled technician reconstructed circuit diagrams from memory. Recall showed marked "chunking", or grouping, by functional units similar to Chess Masters' recall of chess positions. In the second experiment skilled technicians were able to recall more than were novice subjects following a brief exposure of the drawings. This advantage did not hold for randomly arranged symbols. In the third experiment the size of chunks retrieved systematically increased with additional study time. Supplementary analyses suggested that the chunking by skilled subjects was not an artifact of spatial proximity and chunk statistics, and that severe constraints are placed on any explanation of the data based on guessing. It is proposed that skilled subjects identify the conceptual category for an entire drawing, and retrieve elements using a generate-and-test process.

The skill of reading nonverbal, symbolic drawings is important for a wide range of occupations including electronics, engineering, chemistry, and architecture. A skilled electronics technician, for example, must be able to understand complex configurations of symbolic circuit elements, and relate these configurations to hardware in need of repair and to the requirements of circuit design problems. While it seems clear that people become more proficient at this sort of skill as a result of training and relevant experience, it is by no means certain how the improvement should be characterized at the level of mental processes. The present studies were carried out to obtain a more precise description of what it is that people who are skilled at reading symbolic drawings actually know. This kind of description may be useful for later applications aimed at assessing the skill level of people, developing job aids for skilled performance, or improving training in such skills.

The previous research most relevant to understanding the skill of reading symbolic drawings concerns the striking effects of experience on the recall of chess and Go positions. DeGroot (1966), for example, showed that a Chess Master could correctly replace from memory 91% of the chess pieces in a midgame position (approximately 22 of 24 pieces) after only a 5-sec study of the board. A "weak" player could replace only 41% of the pieces correctly. These results have been replicated by Chase and Simon (1973) and others, although the

absolute level of recall by Masters is typically a bit lower. Similar results were found for Go players of different skill levels (Reitman, 1976). The latter two studies also demonstrated that the highly skilled subjects did not have generally superior visual memories, because they had no advantage in recalling random arrangements of pieces.

The "perceptual chunking hypothesis" has been proposed as an explanation of the skill acquired by Chess and Go Masters. It has been hypothesized that Chess Masters, for example, perceive by coding the position of entire chunks, or groups, of several pieces. In one version of the theory (Simon & Gilmarin, 1973), these chunks have verbal labels that can be stored in short-term memory and decoded at the time of recall. It is argued that Masters can quickly represent an entire chess position by a relatively small number of chunk labels that can be used to reconstruct the positions. Players of lesser skill have learned fewer chunks and thus perceive and recall less efficiently. Pauses between successively recalled pieces, the estimated size and number of chunks, and the correspondence of recall groupings with groupings in copying tasks are data cited in support of the perceptual chunking hypothesis.

The first objective of the present studies is to generalize the chess and Go results to the skill of reading symbolic drawings. The question is whether subjects who are skilled at this sort of reading show evidence of using larger chunks than unskilled subjects. Schematic drawings of electronic circuits composed of standard logic symbols were the materials used.

At least on the surface, these drawings share some of the characteristics of the previously studied board games. They consist of symbolic visual patterns in which groups of symbols (e.g., amplifier configurations, rectifiers, filters, etc.) recur frequently, but where great differences exist among entire circuits. On the other

Requests for reprints should be sent to Dennis E. Egan, Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974. We gratefully acknowledge the assistance of Jacqueline D. Cox, who helped run subjects and analyze data for the third experiment. We also thank E. Z. Rothkopf for his thoughtful review of earlier drafts of the paper. A paper summarizing the first two experiments was presented at the annual meeting of the American Educational Research Association, March 27-31, 1978, Toronto, Canada.

hand, circuit drawings differ in two potentially significant ways from chess or Go positions. For one thing, the association between the spatial arrangement of symbols and the functional relation between symbols is not as direct as in the board games. A second obvious difference is that these drawings are not taken from a competitive situation in which moves are planned, pieces are exchanged, and positions progressively develop. The import of these differences is unclear. Certainly the skill of reading symbolic drawings is valuable in its own right, so it is interesting to investigate the degree to which the chess and Go results can be generalized to this practical case.

The second objective of these studies is to contribute to the theory of skilled perception. This was done by observing how subjects' skill level interacts with other factors and by examining the effect of increasing the time to study symbolic drawings.

Three experiments were performed. The first experiment was an intensive investigation of a single skilled subject who had previously indicated his organization of the drawings. This experiment explored several indices of chunking. The second experiment was a comparison of subjects of different skill levels and included two types of controls. The third experiment investigated how recall is affected by increasing the time to study the drawings.

EXPERIMENT 1

Previous studies of skilled perception have attempted to identify functional units or chunks by examining pauses in recall sequences. As Reitman (1976) showed, however, an organizational structure cannot be inferred from such data. In particular, the organizational units of a subject may overlap. Without knowing the organization prior to analyzing the data, there is no way to determine the boundaries of a subject's chunks.

In the first study, an organizational description of each drawing was obtained from a highly skilled subject 3 months before he participated in the recall task. It was assumed that the subject's mental organization of a drawing would be about the same every time the drawing was presented. Under this assumption, transitions among organizational units in the recall sequence can be identified by referring to the previously obtained descriptions. The characteristics (interresponse time, transitional error probabilities) of different types of transitions can be compared directly in this fashion.

Three indices of chunking were investigated. First, it was hypothesized that recall of successive symbols within the same functional unit should be faster than recall of successive symbols from different functional units. The rationale according to the perceptual chunking hypothesis is that crossing a chunk boundary in recall should require additional time to retrieve a new chunk from short-term memory. Second, for similar

reasons, the transitional error probability of within-chunk transitions should be lower than that for between-chunk transitions. Given that symbol n in the recall sequence is correct, symbol $n + 1$ should be an error more often when the n to $n + 1$ transition crosses a chunk boundary, that is, requires initiating a new chunk. Finally, it has been demonstrated that Chess Masters output larger chunks of symbols followed by smaller chunks (Chase & Simon, 1973). The sequential output of groups of diminishing size is also characteristic of subjects recalling conceptually categorized lists of words (Bousfield, 1953). The size of successively recalled groups was therefore taken as a third index of chunking. It was expected that the subject would recall larger groups of symbols followed by smaller groups.

Method

Subject. The subject was a skilled electronics technician at Bell Laboratories. He had over 25 years of experience working with electronic circuits of various types.

Organizational descriptions. Approximately 3 months prior to participating in the recall task, the subject was given copies of 50 electronic circuit drawings. These copies were taken directly from the texts in which they appeared and included figure captions and supplementary information about the function of the circuit. The subject was instructed to indicate the meaningful groups of symbols in each drawing by circling the symbols that, taken together, served a common function. He was also asked to supply a verbal label for each group. The subject was encouraged to give a detailed description and was told that the functional units were allowed to overlap or be nested depending on how he understood the circuit. Nothing was said about the recall task. An example of one circuit and the resulting organizational description is given in Figure 1.

Drawings. The drawings were 50 transistor circuits taken from several texts and circuit design books. Drawings were initially selected so that their function could be easily understood by technicians involved in design of electronic laboratory equipment. From the original 50 drawings, a subset of 36 drawings was selected. These included six representatives of each cell of a 3 by 2 matrix. One factor of this matrix was the type of circuit (multivibrator, power supply, inverter/converter). The

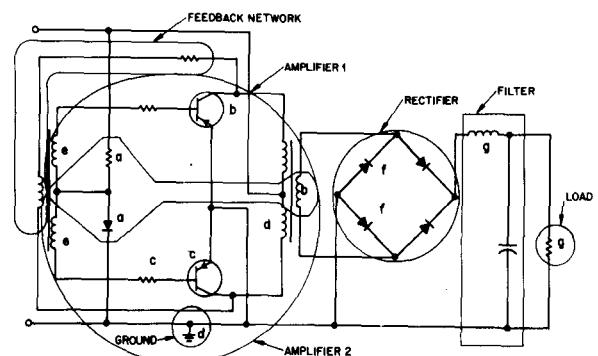


Figure 1. An example of a circuit drawing used in the studies. Enclosed groups and verbal labels indicate the organizational description provided by a skilled technician. Transitions between symbols with the following labels illustrate (a) within-overlap, (b) between-overlap and chunk, (c) within-subchunk, (d) between-subchunk and chunk, (e) mirror-image, (f) within-chunk, and (g) between-chunk transitions.

other factor was the size of the circuit (small, large). Small drawings consisted of 11 to 15 symbols; large drawings had 20 to 29 symbols.

The circuits were redrawn using a standard symbol set—12 common logic symbols (transistor, resistor, capacitor, etc.), each of which occupied a 1.6-cm-sq area. Drawings were laid out to fit into a 25.4 x 15.2 cm rectangle. For each circuit, two drawings were prepared. The first was the entire circuit as it appeared in the text, but stripped of all alphanumeric information. The second was an "answer sheet." Answer sheets consisted of the same pattern of lines or wires as the original circuits, but had blank 1.6-cm squares in place of the circuit symbols.

Recall task. In the recall task the subject was shown each drawing for 10 sec. At the end of this study period, a buzzer sounded and the drawing was removed, exposing the corresponding blank answer sheet. The subject then attempted to reconstruct the drawing from memory, one symbol at a time, by placing magnetized circuit symbols on the blank spaces of the answer sheet. The answer sheet was fixed in the field of view of an overhead camera and adjacent to a digital clock. The experimenter took a picture after each symbol was placed on the answer sheet, thus recording the order of recall and time to the nearest second. The subject had available 15 resistors and 11 of each of the other symbols—more than enough of each type of symbol to reconstruct the drawings. The recall period lasted 90 sec after which the next trial was begun.

Scoring. Each symbol output by the subject was scored correct if it was placed in a correct position on the answer sheet. Transitions between successively recalled locations on the answer sheet were classified in one of seven different categories by referring to the subject's organizational descriptions obtained earlier. The categories were as follows: (a) within-overlap transitions occurred between symbols in an area where two or more functional groups (chunks) overlapped; (b) between-overlap transitions occurred between a symbol in an overlap area and another symbol in the same chunk but outside the overlap area; (c) within-subchunk transitions occurred between two symbols in a subchunk (i.e., a chunk totally nested within another chunk); (d) between-subchunk transitions occurred between a symbol in a subchunk and another symbol outside the subchunk but still inside the chunk; (e) mirror-image transitions occurred between pairs of symbols in different chunks, where the symbols not only served a similar function but were also placed symmetrically about one of the major axes of the drawing; (f) within-chunk transitions were "pure" cases of transitions occurring between two symbols in a chunk; (g) between-chunk transitions were "pure" cases of transitions occurring between two symbols in different chunks. If the subject placed a symbol and later changed it, both responses were scored in the order in which they occurred for the sequential analyses. For analyses of overall number correct, however, only the final response for each location was counted.

Results

In the 90 sec allowed for recall, the subject averaged 10.4 symbols (58%) correct, 2.4 symbols (13%) wrong, and 5.0 (28%) of the blanks unfilled. For this set of drawings, neither the size nor the type, nor the Size by Type interaction, had a significant effect on the mean number of symbols recalled (all F_s had $p > .25$). For the remaining analyses, results were pooled across these factors.

Each of the correct responses (unless it was the first response on a trial) was classified by transition type, and the characteristics of each type of transition are given in Table 1. The first two columns of Table 1

Table 1
Recall Data for Skilled Technician

Transition Type	N	P	P_{max}	IRT	TEP
a. Within Overlap	6	.02	.46	4.5	*
b. Chunk↔Overlap	16	.05	.22	5.2	*
c. Within Subchunk	30	.09	.42	4.5	*
d. Chunk↔Subchunk	9	.03	.13	4.4	*
e. Mirror Image	53	.15	.70	4.6	.00
f. Within Chunk	106	.31	.40	5.0	.11
g. Between Chunk	124	.36	.27	5.8	.21

Note— N = number correct, P = proportion of all responses, P_{max} = proportion of maximum possible, IRT = interresponse time in seconds, TEP = transitional error probability. * N_s were too small to reliably estimate transitional error probabilities for these transitions.

give the number and proportional representation of each type of correct transition. Since transitions are not equally represented in the drawings, the third column adjusts for this by giving the ratio of actual transitions attempted to the maximum number of that type possible.

A greater proportion of the possible transitions within an organizational unit were attempted (.46 of the within-overlap, .42 of the within-subchunk, and .40 of the within-chunk transitions), as compared to transitions between units (.22 of the between-overlap, .13 of the between-subchunk, .27 of the between-chunk transitions). The subject utilized 70% of the maximum possible transitions within mirror-image pairs. Since the transitional error probability for these transitions was zero, symmetry of a drawing can be viewed as an important memory aid. As some subsequent analyses will be restricted to "pure" cases of within- vs. between-chunk transitions, it also should be noted that within- and between-chunk transitions represented 67% of all transitions.

The fourth column in Table 1 gives the average interresponse time (IRT) for each type of transition. In all analyses of interresponse times, only those IRTs preceding a correct response were used, and all IRTs ≥ 15 sec (approximately 3% of the total in this experiment) were discarded. The resulting mean IRT for within-chunk (Type f) transitions (5.0 sec) was faster than that for between-chunk (Type g) transitions (5.8 sec) [$t(225) = 2.34, p < .05$]. A similar pattern was true of transitional error probabilities. The probability of an error in recalling symbol $n + 1$, given that n was correct, was .21 if the n to $n + 1$ transition crossed a chunk boundary. If the transition was within the same chunk, the transitional error probability was .11. These two proportions differ at a marginal level of significance ($\chi^2(1)$ corrected for discontinuity = 3.64, $.10 > p > .05$).

The prediction that larger chunks should be followed by smaller chunks in output is evaluated in Table 2. For this analysis each group of recalled symbols separated by a between-chunk transition (Type g) was classified as a "recall chunk," and the mean size of

Table 2
Number of Symbols per Recall Chunk Output
by Skilled Technician

	Output Position					
	1	2	3	4	5	6
Mean	3.19	2.64	2.45	1.83	2.36	1.82
SE	.34	.32	.32	.21	.36	.20
N	36	36	33	29	22	17

successive recall chunks was computed. Since each between-chunk transition initiated a new recall chunk, recalling part of a chunk and later returning to it counted as two recall chunks. Correct and wrong responses were used since the transition type depended only on the location the subject attempted to recall. The size of recall chunks so defined declined from 3.19 symbols (average size of first group) to 1.82 symbols (average size of sixth group). The subject output an average of 5.4 recall chunks per trial.

Discussion

The IRTs and transitional error probabilities of within- and between-chunk transitions were consistent with expectations, as was the sequence of recall chunk sizes. The types of transitions selected further suggest that the subject attempted to recall systematically by organizational units rather than haphazardly. There is thus support for the claim that the subject's recall of a drawing following a brief study was organized by the functional groups he had identified previously.

EXPERIMENT 2

The second experiment served several purposes. First, it was an attempt to generalize the results of the first study to other skilled subjects who had not worked with these drawings explicitly. Since the circuits are well understood, it seemed likely that other skilled technicians would organize the circuits in a way similar to the organizational descriptions obtained in the first study. Therefore, the previously obtained descriptions also were used to score recall in this study. Second, the question of whether skilled subjects have an advantage over unskilled subjects in recalling symbolic drawings was explored. Third, two control tasks were used. The random-recall task was similar to other such tasks employed in the chess and Go studies. It controlled for the possibility that skilled subjects' advantage in recalling meaningful drawings might be due to superior memory for individual symbols. The construction task was a control for superior guessing. If skilled subjects are simply better guessers, then they should have a large advantage when the task is to guess symbols given the answer sheet alone. Fourth, this experiment was an initial investigation of the effect of study time, although that effect was more thoroughly investigated in Experiment 3.

Method

Subjects. The six skilled subjects were electronics technicians employed at Bell Laboratories. The least experienced subject had been employed 6 years, the most experienced over 20 years in jobs requiring work with electronic circuits. The six novice subjects were college students or lab technicians who had little knowledge of electronics. A typical novice subject could identify some of the circuit symbols by name, but none of the novices could identify all the symbols. By contrast, not only could all the technicians identify all the symbols correctly, but they were also able to easily identify configurations of symbols (trigger networks, rectifiers, etc.) by verbal labels.

Drawings. In addition to the drawings used in the first study, a set of random drawings was prepared. Each random drawing had the same wiring pattern and the same circuit symbols as its meaningful counterpart, but the symbols were positioned randomly in the spaces. Any randomization resulting in 40% or more of the symbols in the same location as the meaningful drawing was replaced.

Recall and construction tasks. Meaningful and random drawings were displayed and recalled by the same method as that used in Experiment 1. Half the subjects were given 5 sec of study, half were given 10 sec. In the construction task, the subject was given no information until the buzzer sounded and the answer sheet was exposed. At that time, the subject attempted to guess the correct symbols, placing them on the answer sheet, one at a time, as in the recall tasks.

Design. Following three practice trials of meaningful recall, and further instructions explaining the random-recall and construction tasks, each subject participated in 12 meaningful-recall, 12 random-recall, and 12 construction trials. The 12 drawings for each task consisted of 2 drawings selected randomly without replacement from each cell of the size (small, large) by type (multivibrator, power supply, inverter/converter) matrix. The 36 experimental trials were randomly ordered.

Scoring. In addition to the criterion used in the first study (position of symbols must be correct), several more stringent criteria were investigated. These ranged from a criterion requiring correct position and orientation to a formula that imposed a penalty for guessing¹ (number strictly correct - .25 × number wrong). Symbols in the random-recall task were scored with reference to their positions in the random drawing. However, *transitions* in all three tasks were scored by referring to the meaningful organizational descriptions. Thus, a within-chunk transition on a random-recall or construction trial involved recalling or guessing successive locations on the answer sheet that, in the meaningful version of the drawing, are in the same functional unit.

Results

Overall performance. Performance by the technicians and novices on each task is shown in Figure 2 for an intermediate criterion (number of symbols placed and oriented correctly). Other criteria produced a similar pattern, with the more stringent criteria (e.g., imposing a penalty for guessing) resulting in slightly larger differences between skill levels. A three-way analysis of variance was performed on subjects' scores, in which skill level and study time were between-subjects factors and tasks was a within-subjects factor. Skill level [$F(1,8) = 6.75, p < .05$] and tasks [$F(2,16) = 44.84, p < .001$] had significant main effects, but study time [$F(1,8) = 2.30, p > .10$] did not. Study time did interact with tasks, however [$F(2,16) = 3.69, p < .05$]. Individual comparisons showed that increased study time aided only in the meaningful-recall task [$t(8) = 2.53, p < .05$].

More importantly, skill level interacted strongly with tasks [$F(2,16) = 9.52, p < .01$], as shown in Figure 2. The technicians performed significantly better than novices on the meaningful-recall task [$t(8) = 2.87, p < .05$] and construction task [$t(8) = 3.81, p < .01$]. The novice subjects performed slightly (nonsignificantly) better on the random-recall task.

Chunking indices. Interresponse times greater than or equal to 15 sec and those preceding a wrong response again were discarded. Using the remaining responses, mean IRTs were computed for each subject on purely within- and purely between-chunk transitions for each task. An analysis of variance was performed on the mean IRTs. Skill level and study time were between-subjects factors; tasks and transition type (within- or between-chunk) were within-subjects factors. The mean IRT for within-chunk transitions (3.9 sec) was more than 1 sec faster than that for between-chunk transitions (5.0 sec). This difference was highly reliable [$F(1,8) = 32.82, p < .001$], but no other effects approached significance. Surprisingly, the difference in IRTs for within- and between-chunk transitions was about the same regardless of skill level, study times, or tasks.

Transitional error probabilities were analyzed using the same design employed for IRTs. The results were rather complex. Much of their import is contained in the interaction of skill level, tasks, and transition type [$F(2,16) = 29.06, p < .001$] given in Table 3. Recalling the definition of transitional error probability [$p(\text{error on } n + 1 | \text{correct on } n)$], large proportions shown in Table 3 reflect a lack of "connectedness" in subjects' output. Skilled subjects' output was generally more connected [$F(1,8) = 20.88, p < .005$], within-chunk transitions showed greater connectedness than between-chunk transitions [$F(1,8) = 1537.25, p < .001$], and tasks were ordered in terms of connectedness of output as follows: meaningful recall > random recall > construction [$F(2,16) = 70.72, p < .001$].

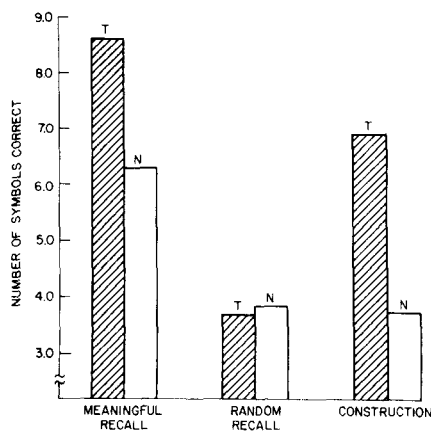


Figure 2. Performance by technicians (T) and novices (N) on meaningful-recall, random-recall, and construction tasks: MSe (between subjects) = 4.10, MSe (within subjects) = .92.

Table 3
Skill Level by Task by Transition Type Interaction for Transitional Error Probabilities

Task	Novices		Technicians		Mean
	Within-Chunk	Between-Chunk	Within-Chunk	Between-Chunk	
Meaningful Recall	.18	.29	.14	.24	.21
Random Recall	.31	.43	.29	.39	.36
Construction	.26	.64	.29	.50	.42
Mean	.25	.45	.24	.38	.33

Note— MSe (between subjects) = .0015; MSe (pooled within subjects) = .0017.

Referring to construction performance in Table 3, if the previous response was correct, skilled subjects made an error 50% of the time when starting to construct a new functional unit. Unskilled subjects erred 64% of the time in the same situation. This is the only individual comparison across skill level that reached statistical significance in Table 3 [$t(8) = 5.86, p < .01$]. This finding suggests that the skilled subjects profit from knowing how two or more chunks combine even when they have not seen the complete drawing. Finally, when interpreting Table 3, note that the transitional error probabilities do not directly reflect the total number of correct transitions. For example, the skilled subjects averaged more correct within- and between-chunk transitions than did novices in meaningful recall, even though the transitional error probability for the two groups was about the same.

Subjects' recall chunks were analyzed as in the first study, and the results are shown in Figure 3. An analysis of variance was performed on the sizes of successive recall chunks in which skill level and study time were between-subjects factors, while tasks and output position (first through sixth chunk) were within-subjects factors. Skill level and output position interacted [$F(5,40) = 3.58, p < .01$] in a way similar to Chase and Simon's (1973) results with chess players. Individual comparisons showed that the first two chunks output by skilled subjects were larger than those of novices. Additionally, tasks [$F(2,16) = 19.16, p < .001$] and output position [$F(5,40) = 13.50, p < .001$] had large main effects. The Skill Level by Task interaction was the only other effect that approached significance [$F(2,16) = 3.03, .10 > p > .05$].

Discussion

Results of the first study generalized to the skilled subjects used in this experiment, even though these subjects had no explicit previous exposure to the drawings. Just as skilled chess players have an advantage over novices in recalling meaningful, but not random, board positions, so it is with technicians and novices recalling symbolic drawings. It therefore seems unlikely that the skilled technicians' advantage is due to superior memory for individual symbols. On the basis of the construction

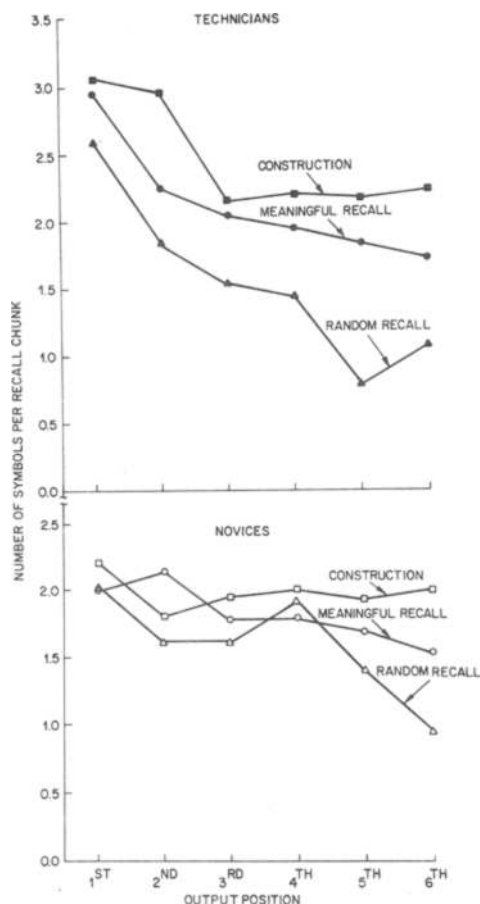


Figure 3. Recall chunk sizes as a function of position of chunk in output in each task for technicians (top panel) and novices (bottom panel). MSe (between subjects) = .65, MSe (pooled within subjects) = .30.

results, however, the idea that their advantage is due to some kind of guessing cannot be rejected. The guessing issue will be developed further in the General Discussion.

The chunking indices often failed to produce interactions that might be expected. While within- and between-chunk IRTs differed in the expected direction, the size of the difference did not depend on the skill level of the subject or the task. The pattern of within- and between-chunk transitional error probabilities was also similar for subjects of both skill levels on the two recall tasks. On the construction task, however, the skilled subjects had an advantage in relating different chunks together. Finally, output position and skill level interacted in the expected direction in the analysis of recall chunks, demonstrating that skilled subjects initially output larger chunks than did unskilled subjects. The surprising aspect of the data relating chunk size to output order was that the pattern did not depend on the task. Technicians used the structure of the drawings in every task, while novices never used it. Thus, technicians attempted to retrieve symbols systematically within the boundaries of a functional unit, even when the symbols were inappropriately positioned (random recall)

or never shown (construction). These data appear to be consistent with a generate-and-test model of retrieval outlined in the General Discussion.

EXPERIMENT 3

The third experiment extended the analysis of study time, and used a slightly different construction control task. Subjects studied the diagrams (meaningful recall) or answer sheets (construction) for 5, 10, or 15 sec before attempting to recall (construct) the diagrams. The main question was whether an increase in study time would lead to better recall and, if so, how the improvement might be characterized. One possibility is that subjects could add more chunks to memory. Another possibility is that subjects could use the same number of larger chunks.

Method

Subjects. The skilled subjects used in this study were six students majoring in electrical engineering at Stevens Institute of Technology and six high-school students selected on the basis of having taken one or more classes in electronics. The unskilled subjects were six college students and six high-school students selected on the basis of having little knowledge of electronics. Preliminary tests requiring subjects to name symbol configurations indicated that the expertise of the skilled subjects in this study fell between that of the technicians and novices of the second study. The unskilled subjects in this study were the lowest group by this measure.

Drawings and tasks. Since the subjects were generally less experienced in this study, only the small drawings were used. The meaningful-recall task was run as in the first two studies. The construction task was modified in that subjects were allowed to study the answer sheet (instead of being given no information) during the allotted study time.

Procedure. Following three practice trials, each subject participated in 12 experimental trials. These were blocked into three groups of four trials having the same study time (5, 10, or 15 sec). Within each block, the subject had two meaningful-recall and two construction trials. On half the trials of each type, the subject worked with a randomly chosen multivibrator circuit, on the other half, a randomly chosen inverter/converter circuit. Trials were randomly ordered within a block of study time, and block sequences were balanced so that each skill group of 12 subjects represented two replications of the six possible block orders.

Results

Overall performance. Preliminary analyses of performance revealed no difference between high-school and college subjects within a skill level and no differences due to the order of study time blocks; results were therefore pooled across these factors. An analysis of variance was performed on the average number of symbols correctly placed (counting orientation). Skill level was a between-subjects factor and tasks and study time were within-subjects factors. Results can be summarized easily (see Figure 4). The skilled students were better on both tasks [$F(1,22) = 19.78, p < .001$], and all subjects were better on meaningful recall as compared to construction [$F(1,22) = 113.92, p < .001$]. Comparing the performance of subjects on this subset

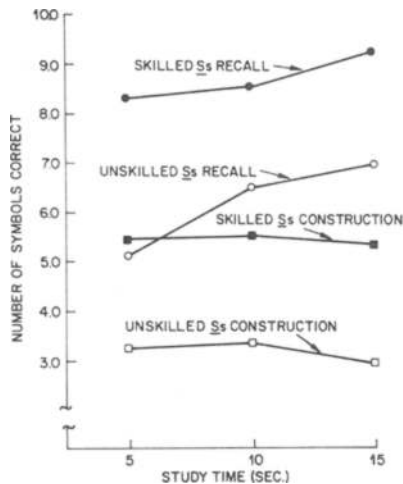


Figure 4. Recall and construction performance by students skilled or unskilled in electronics for different amounts of study time. MSe (between subjects) = 10.57, MSe (pooled within subjects) = 2.34.

of diagrams across Experiments 2 and 3 showed the expected order for both construction and meaningful recall: technicians (Experiment 2) > skilled students (Experiment 3) > novices (Experiment 2) > unskilled students (Experiment 3).

Effects of study time. Additional study time again aided only in recall, the interaction of Task by Study Time (see Figure 4) being statistically significant [$F(2,44) = 5.03$, $p < .05$]. Individual comparisons showed that 15 sec of study led to significantly greater recall than did 5 sec [$t(22) = 3.31$, $p < .01$].

The question specifically addressed in this experiment concerned the nature of the effect of increased study time. More study time resulted in small but significant increases in the size of recall chunks [$F(2,44) = 4.28$, $p < .05$], from 5 sec (where average size of the first six recall chunks was 1.9 symbols) to 10 sec (2.1 symbols) to 15 sec (2.2 symbols). On the other hand, the average number of recall chunks showed no significant effects due to study time [$F(2,44) < 1.0$].

Chunking indices. Because of poor performance of the unskilled subjects and the small size and number of diagrams, transitional error probabilities could not be estimated reliably. Reliable IRTs and recall chunk sizes were available for unskilled subjects only in meaningful recall.

The IRTs for meaningful recall were analyzed separately. The IRTs for within-chunk transitions were again more than 1 sec faster than those for between-chunk transitions [$F(1,22) = 25.53$, $p < .001$]. This time, however, skill level interacted with the within- and between-chunk factor [$F(1,22) = 5.14$, $p < .05$]. For correct within-chunk transitions, the skilled subjects (4.3 sec) took about as much time as the unskilled subjects (4.3 sec). For correct between-chunk transitions, the skilled subjects (4.9 sec) were faster than

the unskilled subjects (6.2 sec), the difference being marginally significant [$t(22) = 1.81$, $.10 > p > .05$]. The meaningful-recall data from the second study were checked and showed the same pattern, but this interaction in Experiment 2 was not statistically significant. The tendency for skilled subjects to be faster than unskilled subjects on between-chunk transitions suggests again that the skilled subjects organize entire chunks together.

The effect of output position on recall chunk size followed the pattern of the second study. Output position once again had a strong main effect [$F(5,110) = 16.06$, $p < .001$]. Although the Skill Level by Output Position interaction did not approach significance in this study [$F(5,110) < 1.00$], the pattern was similar to that obtained before. Skilled subjects' chunk sizes were initially high (average size of first recall chunk was 3.12 symbols) and declined (average size of sixth recall chunk was 1.35 symbols), while the unskilled subjects showed a smaller range of values (first chunk size was 2.54, sixth was 1.42 symbols).

GENERAL DISCUSSION

These experiments successfully extended the findings of other studies of skilled perception to the skill of reading symbolic drawings. This was true despite some potentially great differences between symbolic drawings and the previously studied board games. In fact, recall of circuit drawings by skilled technicians whether measured by number correct, chunk size, or number of chunks, is remarkably similar to recall of chess positions by chess players in the Class A to Master range (cf. Chase & Simon, 1973).

Further discussion will center on three questions. The first is whether the observed chunking is an artifact of spatial proximity and chunk statistics. The second question is whether guessing can account for the advantage skilled subjects have in recalling meaningful patterns. The third question concerns the cause of chunking in recall.

Is Chunking an Artifact?

A possible explanation of skilled perception data is that skilled subjects simply retrieve elements that are close together in a display. In this view, chunking by functional units is a second-order phenomenon due to the spatial segregation of meaningful groups in most organized displays.

Perhaps the strongest evidence for chunking is the pattern of recall chunks in which skilled subjects initially retrieve larger groups of symbols than do unskilled subjects. Even that result can be explained without resorting to chunking. Suppose (a) all subjects are likely by chance alone to start recalling elements within the largest functional unit, (b) all subjects recall as many spatially proximal elements as possible and then

make a transition to another part of the drawing (probabilistically an element in the second largest chunk), and (c) skilled subjects recall greater detail in any given neighborhood. If this model is correct, then scoring output using nonfunctional groups that preserve spatial proximity and chunk statistics should produce the same interaction of skill level and output position as that found by scoring with the functional groups. To test this idea, the following supplementary analysis was performed.

Method. For each drawing a bogus organization was created. The bogus organization was designed to match the skilled subject's original organizational description on the number, size, overlap, and nesting of chunks. Bogus chunks were composed of spatially proximal symbols, but, within this and the other constraints, symbols within a bogus chunk were drawn from as many different real chunks as possible (see Figure 5). The bogus chunks were then used to rescore the data from the second experiment.

Results. Sizes of chunks determined by the real and bogus organizations are given in Table 4. The data shown are the means of chunk sizes averaged across the meaningful-recall, random-recall, and construction tasks for experts and novices. As noted previously, the interaction of Skill Level by Output Position was significant [$F(5,40) = 3.58, p < .01$] when the real chunks were used to score performance. These data are presented in the top two rows of Table 4. When the bogus chunks were used to rescore the data (bottom two rows), the interaction of Skill Level by Output Position vanished [$F(5,40) < 1.00$]. By comparing the second and fourth rows of the table, the close correspondence between real and bogus recall chunk sizes for novices can be observed. The fact that these values are so close, particularly in the first three output positions, suggests that the novices were organizing output on the basis of spatial proximity alone.

The effect of output position on recall chunk size clearly depends on more than the frequency statistics of groups of elements in a drawing. Although this analysis does not totally rule out spatial proximity as

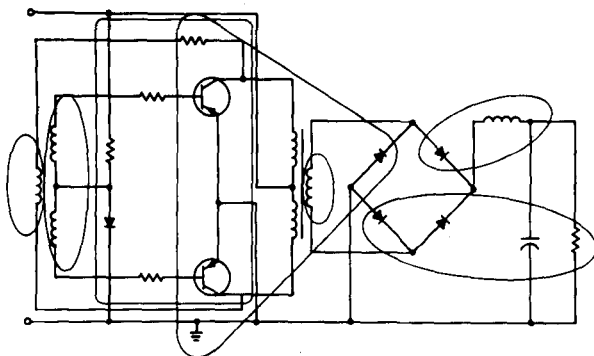


Figure 5. Example of bogus organization corresponding to Figure 1.

Table 4
Sizes of Successive Recall Chunks Defined by Real and Bogus Organizations

Scoring Method	Group	Output Positions					
		1	2	3	4	5	6
Real Chunks	Technicians	2.87	2.36	1.92	1.86	1.60	1.68
	Novices	2.08	1.86	1.79	1.91	1.68	1.50
Bogus Chunks	Technicians	1.99	1.84	1.61	1.53	1.64	1.48
	Novices	2.01	1.83	1.74	1.67	1.51	1.31

the basis of grouping in output, it strongly suggests that the functional nature of the real chunks was the determining factor of the larger groupings for skilled subjects. As for the other chunking indices, the effect of transition type on IRTs and transitional error probabilities for *unskilled* subjects was probably due to spatial proximity. In cases where transition type interacted with skill, the interaction was due to unskilled subjects showing a greater difference than skilled subjects between the within-chunk and between-chunks transitions. Apparently, the skilled subjects' knowledge of relationships among entire functional units helped them to link spatially segregated groups of symbols in recall.

Are Skilled Subjects Simply Better Guessers?

The difference between skilled and unskilled subjects on the construction task was about as large as the corresponding difference in meaningful recall. It might be claimed, therefore, that skilled subjects did not perceive or remember diagrams in a different way, but that they were simply better at guessing.

No data in these experiments conclusively rule out a guessing explanation. There is evidence that rejects simple guessing explanations and constrains the type of guessing possible. First, the formula penalizing for guessing produced larger differences in favor of the skilled subjects in meaningful recall. This finding rejects the simple and unlikely idea that skilled subjects guess more often, but with the same accuracy, as unskilled subjects. Second, it might be proposed that skilled subjects recall all they can and then start guessing. In this view, the first part of the recall sequence for skilled and unskilled subjects should be similar, but the latter part should differ in favor of the skilled subjects. This simple model also must be wrong. Even in recalling the first few symbols, the skilled subjects retrieved in a different pattern, using larger functional groups. Third, we can define an entire class of simple guessing theories based on the probability of correctly guessing symbols given no information. These simple guessing theories have difficulty with the very fact that the advantage of skilled subjects is about the same after zero study time (construction) as it is after 5, 10, or 15 sec of study (meaningful recall). Suppose skilled subjects can guess a subset of the symbols, given no information, and their per-

formance on this subset leads to the same advantage after 0 to 15 sec of studying a diagram. Since the performance of all subjects improves with increased study time, any simple guessing theory must claim that skilled subjects use the extra study time to selectively remember (at the same rate as unskilled subjects) only those symbols they could not have guessed. Such a process deserves the title "skilled perception."

The most promising guessing theories are not simple, but rather rest on skilled subjects' ability to work with the conditional probability of a symbol given other adjacent information. A paradigm that involves lengthy periods of reconstructive activity is not a good procedure to use in studying these sophisticated guessing theories. Recognition tasks with forced-choice alternatives of the kind used in word perception experiments may be better procedures (see Johnston, 1978).

What is the Cause of Chunking?

Theoretical accounts of chunking in recall of chess positions (Simon & Chase, 1973; Simon & Gilmarin, 1973) have emphasized *perceptual* chunking. Two critical features of the theories are that chunks are perceived independently and that recall consists of decoding verbal chunk labels held in short-term memory. These features lead to several problems. First, chunk independence ignores the perceived saliency of global characteristics of a display. In the present studies, skilled subjects could not only recall the various parts of a display, they also could characterize an entire display after a few seconds (e.g., "some kind of power supply"). Second, a configuration of symbols may or may not form a chunk, depending on its context. Functional groups cannot be independent if they are in fact functional only in a context of other specific groups. Third, recent studies (Charness, 1974; Frey & Adesman, 1976) employing a Brown-Peterson paradigm have found that tasks greatly interfering with memory for verbal materials have only a minimal effect on recall of chess positions. Thus, the studies found no support for the notion that visual patterns are represented by verbal labels in short-term memory.

An alternative to the perceptual chunking hypothesis links chunking to the organization of concepts in long-term memory. Suppose the following were true. (a) Skilled subjects identify the appropriate conceptual category for an entire display. They may accomplish this by processing global display properties or by coding a single chunk whose relation to a more general category is known (Palmer, 1977). (b) Knowledge of the conceptual category enables skilled subjects to retrieve symbols systematically, because functional units are conceptually related to display categories in long-term memory. For example, skilled technicians know that a power supply is likely to include a source, rectifier, filter, regulator, etc. Without any further effect on the intake of information, knowing the conceptual

category of a display may result in grouping output by its functional units. (c) Knowledge of the conceptual category may also enable skilled subjects to search drawings more systematically. This leaves open the possibility of verifying local details suggested by a global category.

The "conceptual chunking hypothesis" appears to overcome the difficulties of perceptual chunking and is consistent with several aspects of the present results. First, skilled subjects recalled larger chunks than did unskilled subjects. This finding may be a consequence of a generate-and-test process. By knowing that a display contains a certain type of functional unit, skilled subjects can generate more symbols within each unit for a strength or recency check (Bower, Clark, Lesgold, & Winzenz, 1969). Second, skilled subjects attempted to output symbols by functional units even in random recall and construction. This finding suggests that the rather sketchy features of the answer sheets were sufficient to cue skilled subjects to the conceptual category and, hence, the functional units of a drawing. Given knowledge of a functional unit in random recall, skilled subjects had difficulty generating candidate symbols that could pass the strength test. In construction they output easy-to-generate but untestable symbols. The greater chunk size (Figure 3) and higher error rates (Table 3) in construction, as compared to random recall (Experiment 2), appear to be generally consistent with this account. Third, skilled subjects were superior to unskilled subjects on between-chunk transitions. A well-organized conceptual structure can also account for this result. A conceptual category characterizing an entire display would make for an excellent retrieval plan linking together spatially segregated groups of symbols. Finally, the average size, but not the average number, of chunks increased systematically with increased study time. Since this pattern was true for subjects of different skill levels, it is not critical evidence. However, the result is at least consistent with conceptual chunking. If skilled subjects know the conceptual category of a drawing, it seems reasonable that they would flesh out the details of the drawing, rather than remember entirely different parts of it, when given additional study time.

In summary, the findings of the skilled perception studies using chess and Go players can be generalized to the very practical case of technicians reconstructing symbolic drawings from memory. However, the theoretical explanation of those findings probably should be altered somewhat. It may be the case that the truly striking characteristic of Master chess players and skilled technicians is not their perceptual skill at rapidly coding meaningful clusters of several elements. Instead, the skill may be more conceptual in nature, allowing skilled subjects to (a) rapidly identify a concept that characterizes an entire display and relates many element clusters together, (b) systematically retrieve elements by

a generate-and-test process, and (c) search drawings systematically by verifying local details suggested by general display concepts. In order for these theoretical ideas to gain credibility, further research must show that skilled subjects can directly access conceptual display categories without first coding each constituent chunk.

REFERENCES

- BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 1953, 49, 229-240.
- BOWER, G. H., CLARK, M. C., LESGOLD, A. M., & WINZENZ, D. Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 323-343.
- CHARNES, N. *Memory for chess positions: The effect of interference and input modality*. Unpublished doctoral dissertation, Carnegie-Mellon University, 1974.
- CHASE, W. G., & SIMON, H. A. Perception in chess. *Cognitive Psychology*, 1973, 4, 55-81.
- DEGROOT, A. D. Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving: Research, method and theory*. New York: Wiley, 1966.
- FREY, P. W., & ADESMAN, P. Recall memory for visually presented chess positions. *Memory & Cognition*, 1976, 4, 541-547.
- JOHNSTON, J. C. A test of the sophisticated guessing theory of word perception. *Cognitive Psychology*, 1978, 10, 123-153.
- PALMER, S. E. Hierarchical structure in perceptual representation. *Cognitive Psychology*, 1977, 9, 441-474.
- REITMAN, J. Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology*, 1976, 8, 336-356.
- SIMON, H. A., & CHASE, W. G. Skill in chess. *American Scientist*, 1973, 61, 394-403.
- SIMON, H. A., & GILMARTIN, K. A. Simulation of memory for chess positions. *Cognitive Psychology*, 1973, 5, 29-46.

NOTE

1. The correction term, $.25 \times$ number wrong, is based on an estimate of correctly placing a symbol by chance, given knowledge of the overall frequency of different symbol types in this set of drawings.

(Received September 6, 1978;
revision accepted January 16, 1979.)