

Skilled Perception in Go: Deducing Memory Structures from Inter-Response Times

JUDITH S. REITMAN

The University of Michigan

Experts appear able to handle much larger amounts of specialized information than nonexperts, and handle it without an apparent superior memory capacity. This finding, based on research on chess players with chess information, was replicated on Go players with Go information. Assuming this superiority occurs because the experts process chunks of information through their limited capacities rather than individual elements, the question then becomes one of defining what the chunks are and how they are related. To this end, the technique of partitioning recall and reproduction data into chunks on the basis of inter-response times (IRTs) (introduced in their work on chess by Chase and Simon, 1973) was applied to the reproduction and recall of Go patterns by a Go Master and a Go beginner. Unlike its application in chess, no single IRT was able to produce consistent, veridical chunks for either Go player. Subsequent analysis of the underlying assumptions of the technique showed it to be limited to only those patterns that can be partitioned into a linear set of chunks, not nested chunks, and to situations in which retrieval and overt recall of each chunk is completed before retrieval of the next chunk. In a supplementary task, the Master Go player indicated that the Go patterns were not seen as linear chunks nor as strictly nested hierarchies, but rather as overlapping clusters. IRTs were found to be correlated with this structure, but were not reliable enough to reflect its details.

Experts differ from nonexperts more in perceptual-memorial abilities than in logical, problem solving abilities normally thought characteristic of thinking. There is something about the expert's accrued past experience and the similarity of the current situation to that experience that allows him to reduce masses of specialized information, without concomitant loss of detail, into units his limited capacity can handle. The expert seems to be built the same as the nonexpert and have the same mechanisms available to him; he is just better at moving information through his system.

The research was supported by Grant DCR 71-02038 from the National Science Foundation, and the Advanced Research Projects Agency of the Department of Defense, monitored by the Air Force Office of Scientific Research under Contract No. F44620-72-C-0019. I wish to acknowledge the assistance of Linda Springer and Zylphia Orr in data analysis, Bruce Wilcox and James Kerwin whose insights into their performance as subjects were invaluable, and to Susan Chipman, Douglas Medin, and Robert Pachella for their very helpful comments on an early draft. Requests for reprints should be addressed to the author at University of Michigan, Human Performance Center, 330 Packard, Ann Arbor, MI 48104.

Such a characterization of the skilled thinker is based primarily on the research of de Groot (1966), Chase and Simon (1973a, 1973b), and Charness (1974), in their explorations of the skilled thinking involved in playing chess. de Groot, for example, found that the depth and breadth of the Master chess player's search in his potential-move tree is equivalent to that of the beginner; the Master explores only good moves, the beginner explores some bad. Also, the basic abilities to reproduce patterns appear to be equal for the two players when the material tested is non-meaningful patterns of chess pieces. The Master shows remarkable recall, however, when tested on real chess game positions, nearly twice as good as that of the beginner.

The question then arises as to what the skilled thinker sees and remembers. What are the structures he notes in the stimulus configuration that then trigger the appropriate actions? The expert's recall superiority without a capacity superiority or a loss of detail suggests that he processes familiar subpatterns of information, called chunks, rather than individual elements. The question then becomes one of defining what the contents of these chunks are and how they are organized.

What behavioral correlates indicate the structure seen and remembered by a skilled thinker? Previous concerns with clustering in memory have resulted in measures of organization too specialized to be of help in this situation. For example, measures of subjective organization (Tulving, 1962; Mandler, 1967) are little more than numbers reflecting the consistency with which a set of elements is recalled contiguously, reflecting the amount of existing structure rather than its form. Similarly, the transition error probability measure, which defines substructure boundaries as those points in serial recall at which error rates are high (Bower & Winzenz, 1969; Martin & Noreen, 1974), applies only to material that must be recalled in a designated serial order and that is not as yet well learned.

As a behavioral correlate of structure, the timing characteristics of recall appear to be more applicable to the kinds of well-learned patterns found in chess. When elements to be learned are preorganized into chunks and elements of a chunk are retrieved together before being overtly recalled, subjects can be expected to recall elements in bursts, pausing between bursts. Because pauses have been shown to fall at the boundaries of experimenter-defined chunks, (Bower & Springston, 1970; McLean & Gregg, 1967; Gelfand, 1971), pauses can be used to discover unknown chunks.

Pauses in recall were adopted by Chase and Simon (1973a) as indicators of the boundaries of the chunks skilled chess players see in the material they perceive and recall. Chase and Simon partitioned the recall of a chess Master, a Class A player, and a beginner into chunks on the basis of a single inter-response time (IRT). Those elements that were

recalled in less than a fixed amount of elapsed time since the last element were considered members of the same chunk, those that required more were considered arising from two successive chunks.

Chase and Simon's results showed that although the contents of the chunks of the players of different skill levels were similar, the Master's chunks were larger overall, and he recalled more of them after a short study time than the beginner. Because he and the beginner were shown to have the same abilities when trying to recall scrambled chessboards (confirming de Groot's earlier finding), Chase and Simon concluded that in order to remember more chunks than the beginner, the Master must have chunks hierarchically organized, one chunk label serving for recall of two or more subchunks.

The purpose of the present work was to examine in detail the technique of partitioning recall on the basis of IRTs by replicating the work of Chase and Simon on material similar but not identical to chess; namely, the game of Go. In the procedure adopted by Chase and Simon, the subject experienced two tasks, a Perception task, from which the contents of and recall characteristics from chunks are inferred, and a short-term recall task, called the Memory task, which produces the data to be chunked. In both tasks, the subject attempted to reproduce a pattern. In the Perception task, he was allowed to look back at the stimulus pattern as many times as he wished. It was assumed that at each glance back to the stimulus pattern, he perceived and coded one chunk, then turned to the response board and placed on it that chunk's constituent elements. In the Memory task, he tried to recall the entire board pattern after 5 sec of study time.

First, assuming in the Perception task that successive glances bound chunks, Chase and Simon looked at the distribution of IRTs that reflected the recall of elements within chunks and those that reflected the crossing of chunk boundaries. Second, a single IRT was chosen, one that separated the two distributions, as a criterial IRT, below which successive recalls were to be considered from the same chunk, above which they were considered to cross chunk boundaries. The correspondence between elements within chunks defined by glances and those within chunks defined by IRTs lent support to the validity of this technique.

Third, Chase and Simon compared the distributions of IRTs executed in the two tasks. The resultant general correspondence allowed the IRT partition value from the Perception task to be used in partitioning the Memory task data. As a last step in the validation, the contents of chunks in the Memory task were compared with those in the Perception task. Because they matched, conclusions were drawn about the sizes of chunks and the number of chunks recalled by the subjects in the Memory task, reflecting the differences between the Master's and beginner's perceptual structures.

Because of an apparent hierarchical nature of the pieces in Go board configurations, Go was thought to be the ideal source for patterns to be used in a replication and extension of the technique introduced by Chase and Simon. In Go, two players compete to surround maximum territory on a board, a 19×19 grid, with walls consisting of single pieces, called stones. (See Eisenstadt & Kareev, 1975.) Players move in turns, at each turn placing one stone on an unoccupied grid point. Isolated stones are incorporated into higher order units called strings. Two or more stones of the same color immediately adjacent to each other on a grid line constitute a string. Neighboring sets of friendly strings form a group, units of still higher order. It is these higher order units that constitute the walls that surround final territory. Go is thus a game of accretion rather than displacement. The amount of territorial control each player has at any time during the game is a function of the relative security of each player's groups and the placement of the groups over the board.

This paper describes an attempt to use IRTs to partition a Go Master's and Go beginner's recall of meaningful and nonmeaningful patterns of Go stones. A critique of the technique follows from the Master's performance in an additional task and from an examination of the underlying assumptions necessary for deducing structure from pauses.

METHOD

Subjects

Two subjects, a Go beginner and a Go Master, were videotaped while performing two tasks. At the time of the experiment, the Master player was estimated to have played over 7,000 games and was the best non-Oriental Go player in the world, ranked 4-dan.¹ His goal is to become a professionally ranked Go player, and to that end he is currently apprenticed in the house of a Go Master in Japan, now ranked 6-dan, and is successfully competing in tournaments with Go professionals. At the time of the experiment, the beginner had played about 50 games and was assiduously reading introductory Go books and journals. He is the programmer associated with the research group currently producing an intelligent Go playing program (see Reitman & Wilcox, 1975; Reitman, Kerwin, Nado, Reitman, & Wilcox, 1974).

Tasks

The two tasks, called Memory and Perception, were close replications of those constructed by Chase and Simon. In both tasks, the subject was to reproduce patterns of Go stones as quickly and as accurately as possible. On the subject's left was a Go board with one of 28 patterns placed in the lower right quadrant. To the subject's right was an empty Go board and a pile of black and white stones, more than enough of each kind to complete the pattern. In the Perception task, the stimulus pattern remained exposed through-

¹ Go players are ranked on a single ordinal scale ranging in increasing ability from 35-kyu to 1-kyu to 1-dan to 9-dan. Though progress through the high kyu rank is rapid, later increases in ranking are accompanied by much larger amounts of time playing against better players.

out the trial; the subject was allowed as many glances back to it as he thought necessary. In the Memory task, the stimulus pattern was exposed for only 5 sec of study time, then was covered while the subject attempted to reproduce as much of that pattern as he could. If he did not complete the entire pattern correctly on that trial, stones from the response board were removed, and the subject was allowed another 5 sec of study on the stimulus pattern. The study-test sequence was repeated until the subject reproduced the entire pattern correctly.

Materials

The 28 stimulus patterns consisted of two sets of 14, one set for the Perception task, one for the Memory task. In each set, 10 patterns were portions of real Go games, four were nonmeaningful, random clusters of stones. Examples of meaningful and random patterns are shown in Fig. 1. The meaningful patterns, extracted from Masters' games reported in Japanese Go journals, comprised the stones placed in any one quadrant of the board at a quiet point in the game. The nonmeaningful, random patterns were constructed by first determining the average cluster size of stones in a quadrant of real games, then assigning black and white stones randomly to appropriate size clusters, then placing these clusters randomly in the quadrant. As a result, random patterns were as compact as real patterns (widely spaced stones being inherently more difficult to reproduce than connected ones), but consisted of only spuriously meaningful relations. Half of all the stimulus patterns consisted of 22-25 stones, half 12-15, corresponding to the numbers of pieces in the Middle and End chess game positions used by Chase and Simon.

Supplementary Task

Six months after the experiment was completed, the Go Master was given a third task. He was issued a set of paper transcriptions of the 20 real game positions used in the experiment and asked to indicate the partitioning he saw in each pattern. He was asked to circle the individual stones that were related, then, as appropriate, to indicate which groups of stones were related on a higher level. In response to these instructions, he circled some groups of stones on the stimulus patterns and made what he called a list of basic "unitary patterns [which he] perceived and remembered together as one thing [since they had been] illustrated so many times that the various stones phase into a single object." Details of these circled patterns are described in the results as data with which to compare the partitioning based on IRTs from the Chase and Simon technique.

RESULTS AND DISCUSSION

From the videotapes of the subjects' performances were recorded the following data. Each stone placement was defined by a board location and a stone color, and associated with it was the time that elapsed since the last stone placement, called IRT. Notation was made when a stone placement was preceded by a glance to the stimulus board in the Perception task, and in both tasks whether this or the previous placement was made in error.

IRTs were recorded by a single transcriber visually inspecting the videotape, indicating each placement by pressing a reaction-time key input to a computer clock. To measure the reliability of the transcription procedure, IRTs from a sample of three trials were recorded once by each of two

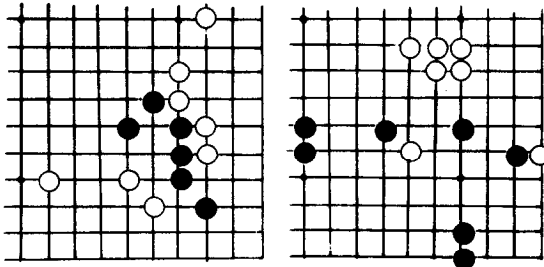


FIG. 1. An example of meaningful (left panel) and random (right panel) patterns.

transcribers. The data from one transcriber were compared point by point with that from the second transcriber. The average differences between these pairs of points was 45 msec; the average *absolute* difference was 133 msec. Since most of the data points were on the order of 1 sec or more, this variability was considered negligible. To measure the veridicality of the transcription, IRTs from another sample of three trials were additionally transcribed by counting frames of the videotape. The human interface procedure matched the IRTs from the frame counting to within 100 msec, again considered negligible.

In what follows, the analytical procedure used by Chase and Simon is traced with the addition of a critique of the procedure based on the Master's performance on the supplementary task. Briefly, both players' overall recall performance in the Memory task reconfirmed the Master's superior recall skills for meaningful but not random patterns. Then, in an effort to describe this superior ability in terms of the numbers of chunks he can retain or the sizes of his stored chunks, a chunk boundary delineator was sought in the IRT data and its validation attempted. The validation procedure involved evaluation of the correspondence between two ways of partitioning the Perception task data into chunks (one which used IRTs and one which used the subject's glances to the stimulus pattern), comparison of the general timing characteristics between the Perception and Memory tasks, a partitioning of the Memory task into chunks on the basis of IRTs, followed by a correspondence check between the contents of chunks from the Memory and Perception tasks.

Overall Retention Performance

Figure 2 illustrates the fact that although the Master reproduced random patterns with the same ability as the beginner, on the meaningful patterns, he was far superior. On Trial 1, the beginner and the Master recalled about 25 and 30% of the stones in the random patterns, respectively. On the meaningful patterns, the Master recalled 66%, the beginner 39%, a significant difference, $t(9) = 3.35, p < .01$. The beginner

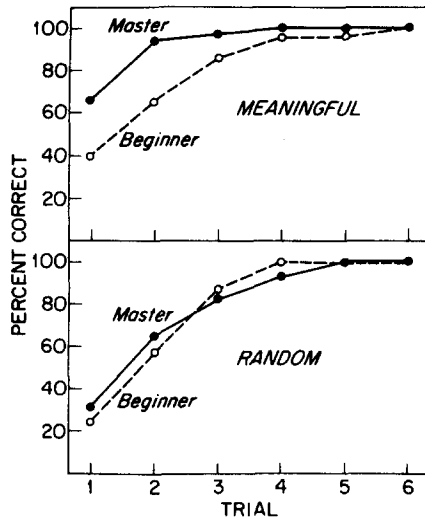


FIG. 2. Percentage recalled by the Master and beginner on successive trials in the Memory task for both random and meaningful patterns.

and Master took on the average 3.75 and 4.25 trials, respectively, to reproduce random patterns; on meaningful patterns, the beginner took 3.4 trials, the Master 2.5, nearly a full trial fewer, $t(9) = 1.86$, $p < .05$.

Search for a Chunk Boundary Delineator in the IRTs of the Perception Task

Whenever two IRT distributions were compared, the Kolmogorov-Smirnov (K-S) two-sample test was applied, as described in Siegel (1956). In all applications of this test, frequency distributions were taken across 100 msec intervals before being converted to cumulative distributions. Results of the applications of this test are reported with an indication of the K-S symbol, the largest difference called D , and the sample sizes.

In the entire set of analyses involving IRTs, only those from errorless performance in meaningful patterns were considered appropriate data. Although the distribution of IRTs for errors in the Perception task was indistinguishable from that of the correct placements, IRTs were significantly longer on the average in the Memory task, K-S: $D(78,90) = .34$, $p < .01$. Also, since the purpose was to define the subjects' structural differences in meaningful patterns, data from meaningful patterns were those upon which the partitioning criteria were validated.

IRTs in the Perception task. The Perception task presumably required the subject to code and retain a chunk each time he looked at the stimulus pattern. Then after he turned to the response board, he placed on

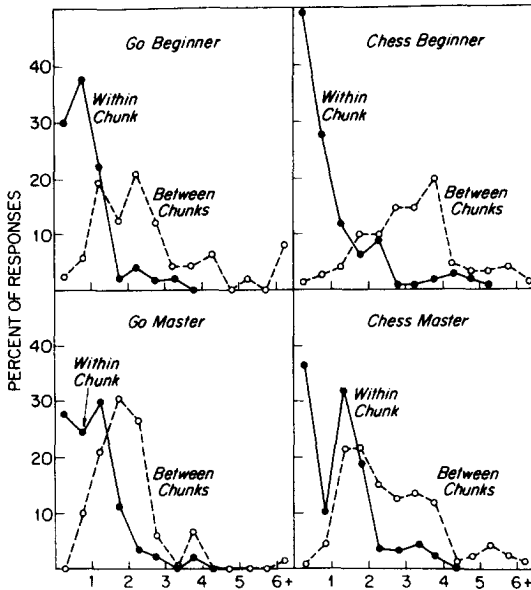


FIG. 3. IRT distributions within and between glance-chunks by beginners and Masters of Go and chess.

the board the stones that constituted the chunk.² The left-hand panels of Figure 3 illustrate the distributions of IRTs that occurred between stone placements that involved a glance back to the pattern (hereafter referred to as "between chunks") and those that occurred without such a glance back to the pattern (called "within chunks"). The right-hand panels illustrate the corresponding distributions for the beginner and the Master chess players from the work of Chase and Simon.

For both Go and chess players, IRTs within a chunk are distinguishably shorter than those between chunks, Go beginner $K-S: D(99,47) = .67$; Go Master $K-S: D(120,42) = .49, p < .01$. The Master players from both games were quicker than the beginners in perceiving the next chunk; the Go Master's IRTs between chunks appear shorter than the beginner's, although significant at only the .10 level, $K-S: D(42,47) = .26$.

The overall timing characteristics were similar for chess and Go except that the IRTs between chunks look somewhat shorter for the Go players than the chess players. Explanation for this probably rests in the kinds of pieces the players have to pick up and place, something they do while glancing. In Go, the pieces are identical except for color; in chess they are different in both color and shape. Picking up the right kinds of

² Inspection of the Master's circled subpatterns, described later, lends support to the notion that subjects in this task do, in general, perceive and place one chunk at a time.

pieces for the next chunk undoubtedly takes longer for chess than it does for Go.

The Go and chess Masters' within- and between-chunk IRT distributions overlapped considerably, more than those of the beginners. For the Go Master, 49% of the within-chunk IRTs overlapped with the between-chunk IRTs; for the chess Master, 58%. The Go and chess beginners' distributions overlapped 45 and 34%, respectively. This overlap is critical to the delineation of chunk boundaries, for it determines the minimum amount of error a single criterial IRT will produce. If the criterial IRT is set at a point at which the numbers of the two kinds of errors are equal, for the Master Go player 21% of the delineated chunk boundaries will be in error. Either stones which are actually part of the previous chunk will be considered spanning chunk boundaries, or the actual beginning of the next chunk will be considered as an extension of the previous chunk. Because of these large overlaps in the IRT distributions, the upper limit of the accuracy that a single IRT may have in partitioning the Master's data is severely reduced.

Contents of chunks in the Perception task. Adoption of the Chase and Simon technique for delineating chunk boundaries on the basis of a single IRT requires validation. Contents of the chunks delineated by IRTs need to be compared with contents of true chunks. Following Chase and Simon, the partitioning provided by glances to the stimulus pattern in the Perception task was assumed to reflect true chunk boundaries. The correspondence between some salient features of these "true" chunks, called glance-chunks, and of the ones created by the IRT partition, called IRT-chunks, was tested.

Stones from within a single chunk could be expected to have one distribution of relationships, and stones that span chunk boundaries to have a different distribution. For example, the stones within chunks are likely to be the same color, forming portions of defense walls or attacking configurations. Similarly, many of the within-chunk stones could be expected to be placed next to each other. Fewer of the between-chunk placements may be expected to be so close; they are more likely to be farther away. Accordingly, in examining the contents of a chunk, five spatial relationships between pairs of successively placed stones were defined and the identity of color between successive stones was noted. The five spatial relationships, illustrated in Fig. 4, were: Next, Diagonal Skip one or two points, Skip one or two on an adjacent axis (similar to knights' moves in chess), and Other.

Whenever in the analyses the kinds of relationships among stones that occurred in one sample were compared with those of another sample, the maximum likelihood estimate chi-square test was used (see Hayes, 1963). In describing where in significantly different distributions the primary differences occurred, $\sqrt{N\phi}$ was used as an indicator of the relative

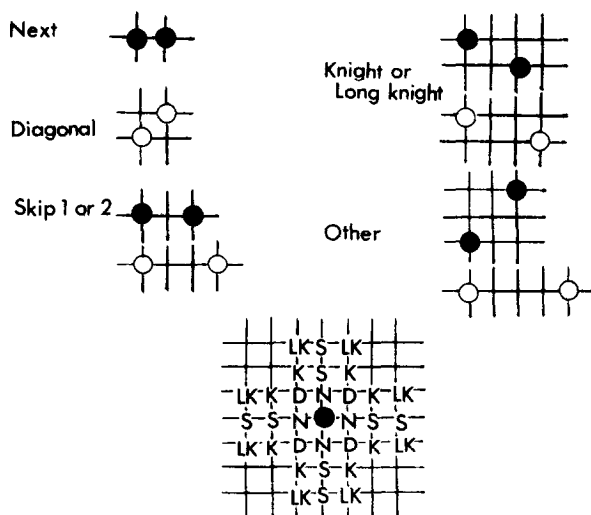


FIG. 4. Exemplars of the five spatial relationships between successively placed pieces and the board territory around a single piece covered by all possible exemplars of these relationships.

magnitudes of the differences (see Haberman, 1973). The $\sqrt{N}\phi$ indicator, calculated from collapsed 2×2 tables from an original $R \times C$ table, approximates the normal deviate, and consequently was of interest (using $\alpha = .05$) when its value exceeded 1.96 in either direction.

Table 1 lists the relationships that existed between successively placed stones both within and between glance-chunks for both subjects. For both the Master and beginner, the distributions of relationships within chunks were significantly different from those between chunks, beginner's $\chi^2(4) = 30.63$; Master's $\chi^2(4) = 11.08, p < .01$. Within chunks, stones were more likely to be placed next to each other, beginner's $\sqrt{N}\phi = +3.03$, Master's $\sqrt{N}\phi = +2.46$; those between chunks were likely to be some "Other" relationship, beginner's $\sqrt{N}\phi = +5.35$; Master's $\sqrt{N}\phi = +2.96$. Similarly, both the beginner and the Master placed more stones of the same color within a chunk than they did between chunks, beginner's $\sqrt{N}\phi = +3.38$; Master's $\sqrt{N}\phi = +3.06$.

Once the distributions of color and spatial relationships that existed within and between glance-chunks were established, the same data were partitioned according to a single IRT criterion and tested: (1) for the same pattern of differences *within* and *between* chunks in color and spatial relationships as found in the glance-chunks, (2) for similarity of the color and spatial relationships *within* IRT-chunks and *within* glance-chunks, and (3) similarity of the relationships that existed *between* IRT-chunks and *between* glance-chunks. Each of these steps required one test for the color relationship, a 2×2 chi-square, and one for the spatial relationships,

TABLE I
 RELATIONSHIPS BETWEEN SUCCESSIVELY PLACED PIECES WITHIN AND
 BETWEEN GLANCE-CHUNKS IN THE PERCEPTION TASK FOR
 BOTH SUBJECTS, WITH PERCENTAGES IN PARENTHESES

	Beginner		Master	
	Within glance-chunk	Between glance-chunks	Within glance-chunk	Between glance-chunks
Same color	80 (81)	23 (49)	98 (82)	24 (57)
Different color	19 (19)	24 (51)	22 (18)	18 (43)
	99	47	120	42
Next	56 (57)	14 (30)	57 (48)	11 (26)
Diagonal	16 (16)	7 (15)	24 (20)	7 (17)
Skip 1, 2	13 (13)	7 (15)	13 (11)	6 (14)
Knight, Long Kn	13 (13)	5 (10)	16 (13)	7 (17)
Other	1 (1)	14 (30)	10 (8)	11 (26)
	99	47	120	42

a 2×5 chi-square. If according to this procedure, the IRT partition did not produce the desired set of similarities and differences, the value of the criterial IRT was changed and the resulting IRT-chunks tested again.

The first criterial IRT evaluated was 2 sec, the one Chase and Simon found acceptable for their chess players. Successive placements that required less than 2 sec were assumed to be within a chunk, those that required more were assumed to cross chunk boundaries. For the beginner, chunks defined by the 2 sec IRT seemed to contain the same kinds of relationships as glance-chunks. He placed more stones of the same color within IRT-chunks than between, $\chi^2(4) = 48.61, p < .01$. And, the color and spatial relationships of placements within glance-chunks matched those within IRT-chunks, color: $\chi^2(1) = 0.09$, relations: $\chi^2(4) = 1.40, p > .10$, and those between glance-chunks matched those between IRT-chunks, color: $\chi^2(1) = 1.26$, relations: $\chi^2(4) = 4.56, p > .10$. The 2 sec IRT boundary misclassified only 10% of the within glance-chunk placements and 39% of the between glance-chunk placements.

Although acceptable for the beginner, the 2 sec criterion did not partition the Master's data well. Although the 2 sec boundary misclassified only 10% of the within glance-chunk placements, it misclassified 60% of the between glance-chunk placements (see Fig. 3). And, although according to most tests the contents of the glance- and IRT-chunks corresponded, the IRT-chunks did not have significantly more same-color stone placements within a chunk than between chunks, as glance-chunks did, $\chi^2(1) = 3.09, p > .10$. The distributions of IRTs suggested that if there was a

single IRT that separated glance-chunks appropriately, it was at a shorter IRT value.

For the Master's data, then, the value of the criterial IRT was reduced in increments of 100 msec until the contents of the IRT-chunks matched that of the glance-chunks. The longest such IRT was 1.3 sec. Chunks partitioned at 1.3 sec contained significantly more stones of the same color within IRT-chunks than between, and the pattern of spatial relationships within IRT-chunks was distinct from that between chunks, color: $\chi^2(1) = 4.82$, relationships: $\chi^2(4) = 18.41$, $p < .01$. The distribution of color and spatial relationships within IRT-chunks matched those within glance-chunks, color: $\chi^2(1) = 0.00$; relationships: $\chi^2(4) = 1.77$, $p > .10$; and those between IRT-chunks matched those between glance-chunks, color: $\chi^2(1) = 0.88$; relationships: $\chi^2(4) = 0.65$, $p > .10$. At this value, the two kinds of misclassification errors were most evenly divided; 34% of the within glance-chunk placements were misclassified as between IRT-chunks, and 30% of those between glance-chunks were misclassified as within IRT-chunks.

General correspondence between Memory and Perception tasks. After criterial IRTs were assigned in the Perception task, the correspondence between the Perception and Memory tasks was assessed. The result of finding these tasks indistinguishable allowed the criterial IRT to be used to partition the Memory task data into chunks.

The Memory task involved scanning the stimulus board for 5 sec, coding chunks of it, and retaining their codes or labels. In responding, the subject presumably took the next code, found its constituent elements in memory, and placed designated stones appropriately on the board. In the Perception task, the subject looked at the stimulus pattern, perceived and coded a single chunk, turned his head to the response board and placed stones appropriately. Implicit in the application of the IRT that was validated in the Perception task to the Memory task are the assumptions that (i) because both tasks involve an identical operation, that of placing on the board stones from within the same coded chunk, the times to execute these moves should be the same in both tasks, and (ii) the operations that are different in the two tasks, turning the head, perceiving and coding the next chunk versus getting a code and finding its constituent parts, require similar if not identical times.

Theoretically, confirmation of only the first assumption was of interest. However, because there was no appropriate definable subportion of the Memory task data, confirmation of both assumptions was attempted simultaneously. A match of the single distribution of IRTs in the Memory task to the combination of the within and between glance-chunk distributions in the Perception task was accepted as support for both assumptions. Figure 5 illustrates the Memory and Perception task IRT distributions for the two subjects. For both the beginner and the Master, the

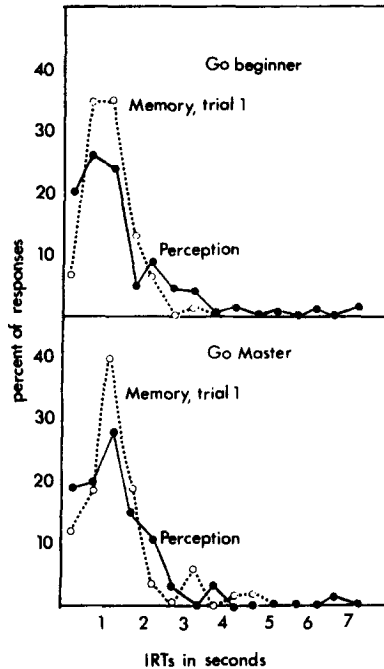


FIG. 5. IRT distributions from the first trial of the Memory task superimposed on the combined within and between glance-chunks distributions from the Perception task for both subjects.

distributions were statistically indistinguishable, beginner K-S: $D(148,54) = .17$; Master K-S: $D(162,89) = .11, p > .10$.

Comparing the contents of Memory task IRT-chunks with that of Perception task glance-chunks. As the last validation step, the contents of the chunks partitioned with IRTs in the Memory task were compared with the contents of the glance-chunks in the Perception task. Table 2 lists the distributions of color and spatial relationships that existed within and between IRT-chunks in Trial 1 of the Memory task. Comparable Perception task data are in Table 1. The IRT used to partition the beginner's data was 2 sec, that for the Master was 1.3.

In the first trial of the Memory task, the beginner made only five responses that required longer than 2 sec. Assuming he did not change strategies drastically, this would indicate that on the initial 5 sec exposure to each pattern, he remembered only slightly more than one chunk. The distribution of relationships involving these five moves was not amenable to statistical comparisons. The kinds of relationships occurring within IRT-chunks, however, could be and were statistically compared with those within glance-chunks. As desired, the distributions of color similarities were identical. However, those involving spatial relationships were

TABLE 2

RELATIONSHIPS BETWEEN SUCCESSIVELY PLACED PIECES WITHIN AND BETWEEN IRT-CHUNKS IN THE FIRST TRIAL OF THE MEMORY TASK FOR BOTH SUBJECTS, PERCENTAGES IN PARENTHESES

	Beginner		Master	
	Within IRT-chunk	Between IRT-chunks	Within IRT-chunk	Between IRT-chunks
Same color	39 (81)	3 (60)	32 (73)	17 (45)
Different color	9 (19)	2 (40)	12 (27)	21 (55)
	48	5	44	38
Next	27 (56)	0 (0)	18 (41)	10 (26)
Diagonal	2 (4)	1 (20)	7 (16)	11 (29)
Skip 1, 2	10 (22)	1 (20)	10 (23)	4 (11)
Knight, Long Kn	5 (10)	0 (0)	6 (13)	8 (21)
Other	4 (8)	3 (60)	3 (7)	5 (13)
	48	5	44	38

Note. The criterial IRT for the beginner is 2 sec, that for the Master is 1.3 sec.

significantly different, $\chi^2(4) = 10.68, p < .05$. Within IRT-chunks in the Memory task, the beginner placed significantly fewer diagonally related stones, $\sqrt{N}\phi = -2.08$, and significantly more "Others," $\sqrt{N}\phi = +2.30$ than within glance-chunks in the Perception task. The chunks were not the same.

For the Master, the single IRT at 1.3 sec was similarly unsuitable. The distribution of spatial relationships within IRT-chunks was not different from that between IRT-chunks, $\chi^2(4) = 6.22, p > .10$. Whereas stones placed within glance-chunks in the Perception task had significantly more "Next" and fewer "Other" relationships, in the Memory task there were only small, nonsignificant deviations in these same directions, $\sqrt{N}\phi = +1.39$ and $-.96$ for "Next" and "Other," respectively.

Perhaps the failure of a single IRT to partition the Memory task data into chunks that match those in the Perception task was due to the selection of the wrong IRT values. There may have been other IRTs that partitioned the Perception task as well as the 2 and 1.3 sec criteria, yet also partitioned the Memory data into chunks with comparable contents.

In search of an acceptable IRT, data from both tasks were partitioned into IRT-chunks using 10 different IRT values, all those between 1 and 2 sec at 100 msec intervals, and subjected to the same kinds of content analyses as described above. Although for both subjects, many IRT values partitioned the Perception data such that the contents of the IRT-

chunks matched that of the glance chunks, no IRT partitioned the Memory task data so that Memory IRT-chunks matched glance-chunks. It appeared, then, that either the two tasks evoked different strategies that resulted in different partitionings or IRTs did not appropriately reflect perceived structure.

That the Master was quick to both perceive the next chunk and to pick up the appropriate response pieces, resulting in the overlap of the between glance-chunk and within glance-chunk IRT distributions, does not seem to be sufficient cause for rejecting this technique for partitioning the data. A single IRT partitioned the Perception task data well. There were enough differences between the moves made in a short time and those requiring a long time that the errors of misclassification did not alter the match between glance-chunks and IRT-chunks within the Perception task. Difficulty came only when the contents of the chunks from the Perception task were compared with those of the Memory task.

Master's Penciled Partitions

Results from the Master's third task, however, suggested a reason why the technique of partitioning chunks with a single IRT failed. In the third task, the Master was given paper transcriptions of the real board stimuli from the Perception and Memory tasks and asked to indicate with penciled circles the partitioning he saw on these patterns. He was expected to circle small subpatterns, then enclose them in higher level patterns, indicating how these groups of stones function together on several levels. The left panel of Fig. 6 illustrates examples of his responses to these instructions. He did not partition these patterns in a strictly nested hierarchy, as expected, but rather circled overlapping subpatterns, some elements of which were entirely enclosed in higher level patterns, some of which were members of two or more chunks. In addition, the Master commented that there was yet another partition of the patterns, that of successive overlapping pairs of stones that reflect the order of actual play of the game, each pair being a move and the opponent's standard reply.

This penciled partitioning, taken as veridical, can aid in assessing the reason IRTs failed to partition the Perception and Memory tasks into equivalent chunks. As mentioned above, one possible reason they failed is that the chunks seen in the two tasks were different, that the two tasks evoked different partitions. Evidence for this conjecture appears through examination of the correspondence between the penciled chunks boundaries, the glances in the Perception task, and the kind of recall in the first trial of the Memory task. If the glances occurred at penciled chunk boundaries and recall consisted primarily of complete penciled chunks then there is little to suggest that the tasks evoked different partitions.

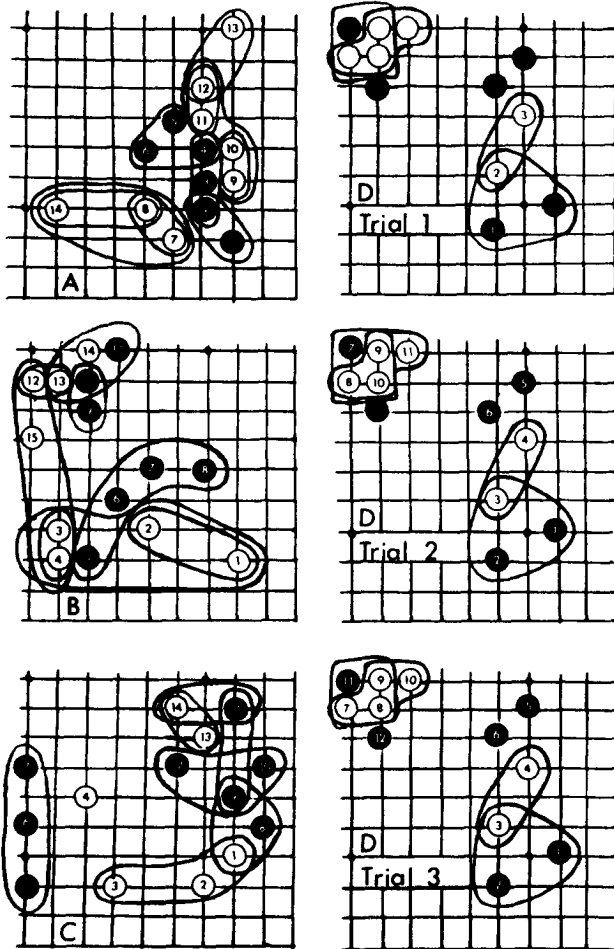


FIG. 6. Examples of the Go Master's penciled partitioning of meaningful patterns and the orders in which he recalled the elements.

Table 3 shows the relationships between the occurrence of glances in the Perception task and the crossing of penciled chunk boundaries. Of the 61 successive placements that crossed penciled chunk boundaries, 33 were accompanied with glances back to the stimulus board; of the 116 placements within a penciled chunk, 96 lacked a glance to the board. Glances appeared appropriately in 78% of the placements. Furthermore, within the 28 cross-boundary placements that were not accompanied with a glance and the 20 glances occurring at places other than penciled boundaries, 14 were within one placement of a glance, either early or late, and six occurred at the point of overlap between penciled chunks. In the cases where the glances were one move off, it is reasonable

TABLE 3
CORRESPONDENCE BETWEEN THE OCCURRENCE OF GLANCES AND
THE CROSSING OF PENCILED CHUNK BOUNDARIES

	Glance taken	Glance <i>not</i> taken	
Partition boundary crossed	33	28	61
Partition boundary <i>not</i> crossed	20	96	116
	53	124	177

to assume that the Master, in being late, initiated the beginning of the next chunk before glancing to the stimulus board, or in being early, glanced while knowing the completion of the preceding chunk. If these 20 almost right responses are included as correct, a total of 84% of the successive placements concur with the glances.

In the first trial of the Memory task, the Master placed at least one stone in each of the 53 distinct penciled chunks, 44 of which were recalled completely (83%). Similarly, of the 115 stones correctly placed, 90% of them were accompanied by all their chunk-mates. It appears that the Master primarily recalled whole chunks, as Tulving and Pearlstone (1966) found, and that the chunks were the same as those indicated in the penciled partitions.

The fact that the penciled partitioning of the patterns in the Perception and Memory tasks correspond to behavior in both tasks, i.e., glances occur at chunk boundaries and recall consists of whole chunks, suggests that the tasks did not force the Master to partition the patterns differently.

It appears then that the mismatch in the chunks delineated by IRTs in the Perception and Memory tasks is not the consequence of differing partitions evoked by the tasks, but rather something more fundamental about IRTs. Suppose it were true that IRTs within chunks were consistently shorter than those between chunks. In the overlapping chunk structure the Master drew, it is not clear what the pattern of IRTs would be. For example, in Fig. 6A, when the Master placed stone 6 then 7, he clearly crossed a chunk boundary; his IRT could be expected to be long. However, in placing stones 1 and 2, he keeps within the first chunk, but simultaneously enters the second chunk (consisting of stones 2, 3, and 4). The placement is both within and between chunks. Are IRTs expected to be long or short?

Figure 7 shows the relationship between the actual IRTs obtained and four definable kinds of successive placements: (1) those clearly within a chunk, (2) those clearly between chunks, (3) those between low-level chunks but nested within a higher level chunk, and (4) those at points at which two chunks overlap. It appears that placements that clearly cross

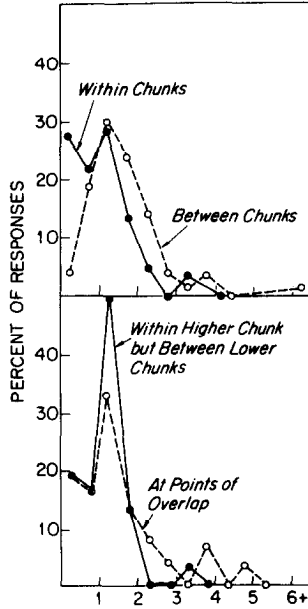


FIG. 7. IRT distributions within and between penciled chunks (top panel) and within nested chunks and at overlap points (bottom panel).

chunk boundaries are longer than those that remain within chunks, but the overlaps among distributions are even more severe than those in Fig. 3. This suggests that although IRTs are related to performance from this kind of overlapping structure, they are not stable enough to be a diagnostic tool.

The General Applicability of the IRT Partitioning Technique

The technique of partitioning recall into chunks on the basis of a single IRT is based on four assumptions. (1) Elements of patterns are stored in memory in chunks, such that each chunk has a label that refers uniquely to its constituent elements or subchunks. (2) If the chunks are part of a hierarchy, they are strictly nested; no one element or subchunk is a member of more than one chunk at a higher level. (3) All elements of a chunk are recalled before all elements of another chunk. And, (4) pauses reflect recall of chunks at one consistent level in the hierarchy.

Because these assumptions are strict, the technique appears to have limited applicability in delineating at least the skilled perception of Go patterns and potentially more. First, if chunks are hierarchically organized and strictly nested, a single IRT chosen to delineate boundaries could be expected to reflect recall of elements at one level of the hierarchy but neither the constituent subchunks nor its higher order suprachunks.

Second, the partitioning of recall on the basis of IRTs requires that all elements of a chunk be recalled before all elements of another chunk. In a situation in which words were presented in an organized set of non-overlapping categories, Gelfand (1971) found that on only half of the trials did subjects show recall that was perfectly categorized. On some trials, subjects would recall most of the elements of a category before moving on to a new category, then return to a previously entered category to attempt completion. On other trials, where there may have been spurious cross-category connections, elements of category may have been recalled with this cross-category item intruding, the second, already entered category being the next to be recalled.

Both of these recall patterns may reflect the kinds of overlapping clusters the Master Go player indicated in his penciled partitions of Go patterns. For example, in Fig. 6B, the Master entered the 3-4-15-12 chunk on two occasions, at the third and fourth item recalled while he was recalling the lower large chunk, and later in completing the upper chunk. For patterns with overlapping chunks, it is often impossible to recall all elements of a chunk before entering another chunk.

Third, the Memory and Perception tasks may have induced chunking at different levels. In the Perception task, where the subject was free to look at the stimulus pattern as often as he liked, he may have encoded chunks at the easiest lowest level. However, in the Memory task, when given only 5 sec of study time on the pattern before being required to recall, he may have opted to encode higher level chunks, looking for far reaching relations that allow more complete recall. Some long IRTs reflect lower level chunks, as in the Perception task, whereas other long IRTs reflect higher level chunks, as in the Memory task. The kinds of relationships among elements within these two levels of chunks could be expected to be different, as found in the analysis of chunk contents.

Finally, the Memory task may have induced the subjects to guess more than the Perception task, with the Master's guessing being more likely to be right than the beginner's. Given some initial stones as clues and the knowledge that the pattern came from a Master level game, the Master may have been able to figure out the final patterns' compositions. Some of his long IRTs may therefore have reflected not the crossing of chunk boundaries but rather problem solving operations that helped him generate good guesses. This criticism has been levied against the chess work before, and in answer the work of de Groot (1966) is cited. de Groot showed that Master chess players are not better than beginners in guessing patterns when they are given as a clue the squares that the pattern occupies. However, the fact that patterns in Go change by growing, rather than by displacement as in chess, may account for the emergence of guessing in this experiment, especially when the subject is given as clues salient pieces (the first ones he sees and remembers) rather than positions

occupied. The possibility of differential guessing in these tasks cannot be easily dismissed.

In sum, it appears that both because IRTs can be affected by the level within a hierarchy a subject encodes and by subject's guessing abilities, and because perceived patterns may not be strictly nested hierarchies, partitioning on the basis of a single IRT is unlikely to produce veridical chunks. Before IRTs can be used as behavioral indices of memory structures, better delineation is needed of: (1) the set of possible structural organizations experts may have, (2) how structures are used in the particular task situations, and (3) tasks that minimize intelligent guessing.

Postscript on a Behavioral Correlate of Structure

As a postscript to the analysis of behavioral correlates of perceived structure, mention should be made of an additional source of systematic performance observed in the Memory task. The numbers indicated on the stones in the left panel of Fig. 6 indicate the order in which the Master placed stones on the board in his final perfect recall of three patterns from the Memory task. The order of stone placements seems to follow, to a large extent, the organization the Master indicated by pencil 6 months after the recall session. Similarly, the right panel of Fig. 6 shows representations of the Master's recall of stones in his three attempts to reproduce a single pattern in the Memory task. Only those stones that were recalled on that trial are numbered. Again, it appears that the recall output orders, though varied, reflect the penciled partitioning. In particular, in Fig. 6D, Trial 2, note that stones 8, 9, 10, and 11 are recalled contiguously within the indicated chunk, and again in Trial 3 as 7, 9, 8, and 10. They are consistently recalled together, without intrusion from elements of other chunks, though their internal order varies. Perhaps this general correspondence between indicated organization and overall consistencies in the recall output order suggest a behavioral correlate to memory structure more applicable to Go patterns and other non-nested hierarchies than IRTs.

REFERENCES

- Bower, G. H., & Springston, F. Pauses as recoding points in letter series. *Journal of Experimental Psychology*, 1970, **83**, 421-430.
- Bower, G. H., & Winzenz, D. Group structure, coding, and memory for digit series. *Journal of Experimental Psychology*, 1969, **80**, (2 Pts. 2).
- Charness, N. *Memory for chess positions: The effects of interference and input modality*. Unpublished doctoral dissertation, Carnegie-Mellon University, 1974, and C.I.P. Working Paper No. 262.
- Chase, W. G., & Simon, H. A. Perception in chess. *Cognitive Psychology*, 1973, **4**, 55-81. (a)
- Chase, W. G., & Simon, H. A. The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973. Pp. 216-281. (b)
- de Groot, A. Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving*. New York: Wiley, 1966.

- Eisenstadt, M., & Kareev, Y. Aspects of human problem solving: The use of internal representations. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition*. San Francisco: W. H. Freeman, 1975.
- Gelfand, H. Organization in free recall learning: Output contiguity and interresponse times as a function of presentation structure. Human Performance Center, University of Michigan, Technical Report No. 29, June, 1971.
- Haberman, S. J. The analysis of residuals in cross-classified tables. *Biometrics*, 1973, **29**, 205-220.
- Hayes, W. L. *Statistics for psychologists*. New York: Holt, Rinehart & Winston, 1963.
- Mandler, G. Organization and memory. In K. W. Spence & J. A. Spence (Eds.), *The psychology of learning and motivation*. New York: Academic Press, 1967. Vol. 1.
- Martin, E., & Noreen, D. L. Serial learning: Identification of subjective sequences. *Cognitive Psychology*, 1974, **6**, 421-435.
- McLean, R. S., & Gregg, L. W. Effects of induced chunking on temporal aspects of serial recitation. *Journal of Experimental Psychology*, 1966, **74**, 455-459.
- Reitman, W., Kerwin, J., Nado, R., Reitman, J., & Wilcox, B. Goals and plans in a program for playing Go. *Proceedings of the 29th National Conference of Association for Computing Machinery*. New York: Association for Computing Machinery, 1974. Pp. 123-127.
- Reitman, W., & Wilcox, B. Perception and representation of spatial relations in a program for playing Go. *Proceedings of the 30th National Conference of Association for Computing Machinery*. New York: Association for Computing Machinery, 1975. Pp. 37-41.
- Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956. Pp. 127-136.
- Tulving, E. Subjective organization in free recall of "unrelated" words. *Psychological Review*, 1962, **69**, 344-354.
- Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 381-391.