SIMULATION OF TASK PERFORMANCE AS GUIDE TO THE IDENTIFICATION OF SOLUTION STRATEGY FROM EYE-MOVEMENT RECORDINGS

Gerhard Deffner

Department of Psychology, University of Hamburg

Von-Melle-Park 5, 2000 Hamburg 20, West Germany

phone: 0049 40 4123 5471

Summary

An approach is presented which was developed for the goal of comparing subjects with respect to their use of solution strategies. They were given a series of eight tasks of a type which can well be solved by two distinctly different methods: n-term series tasks in the form of statements about spatial relations. One of these methods relies on mental imagery (Method of Series Formation) while the other is more of an abstract, analytical method (Elimination). For the purposes of comparison, eye-movements were recorded during task performance. The sequence of gazes recorded during each individual solution attempt was subsequently matched to patterns derived from information processing models of task performance which simulate the different strategies. The percentage of successful matches to one or the other pattern was used in further analyses comparing the two groups of subjects.

Keywords: cognitive simulation, eye-movement data, strategy identification

Identification of Strategies

1. Introduction

The methodology described here was developed for an experiment which continued a series of studies (Deffner, 1984) in which silent control subjects were compared to subjects who were required to think aloud during their attempt to solve experimental tasks. In the new experiment, the main variable of interest was strategy use, especially shifts in the selection of strategies over time.

2. Material

The experimental tasks were derived from Ohlsson's (1980) collection of tasks. They consist of n-term series tasks where propositions refer to the relative positions of persons sitting on a bench, thus the 'terms' in these tasks are names, and the propositions use spatial relations. In contrast to Ohlsson, who for the purposes of studying the solution of more complex tasks used a larger number of relational terms (immediately to the left/right of, left-/rightmost, left-/rightmost but one, left/ right, between) only one type of relation was used in the present context. An example is given in Figure 1, where line numbers are included to aid in the understanding of subsequent examples.

```
SEVERAL  BOYS  ARE  SITTING  ON  A  BENCH:        (line 0)

         KEITH  IS  LEFT  OF  CHUCK               (line 1)
         CHUCK  IS  LEFT  OF  ROY                 (line 2)
         MARK  IS  RIGHT  OF  ROY                 (line 3)
         MARK  IS  LEFT  OF  PHIL                 (line 4)
```

Figure 1: n-term series task

In each task, subjects are required to answer a specific question with respect to the the position of one particular person, for example:

"Who is immediately to the right of Mark?"

Eight experimental tasks were used (four 4-term and four 5-term series tasks). Eye-movements were recorded using an Applied Sciences 1996 system.[1] This equipment measures x- and y-coordinates for gaze direction at a sampling rate of 60

[1] I want to thank Richard Ohlsson from the University of Colorado for the privilege to use his laboratory and for his generous help.

Hz, and recordings are made on the level of fixations (Kliegl & Ohlsson, 1981). This means that successive samples are aggregated as long as they lie on the same character position of the display (Kliegl, 1981). This aggregation is very fine and it results in a resolution of distinct fixation points <u>within</u> individual words. For the present analysis, a more natural unit of analysis is the 'gaze' rather than 'fixation'. Gazes incorporate all fixations on the same word. This resolution was chosen because in the present context we are not concerned with an analysis of the reading process during task performance but want to use units which are relevant entities in information processing models of task performance (for a discussion of this principle, see: Just & Carpenter, 1976). For this reason, only three gaze positions were identified per line:

1) the term (name) on the left,
2) the relational information in the middle ("is to the left/right of"),
3) the term (name) on the right.

Figure 2 shows the beginning of a recorded sequence of gazes where the first number in each parenthesis stands for gaze position relative to the layout of the task (the first digit refers to line number and the second digit stands for first, second or third position in a proposition); the second number stands for the duration of the gaze on that location.

(03 38)(02 7)(01 13)(03 25)(11 5)(13 15)(12 7)(11 7)
(13 70)(21 8)(13 36)(12 7)(11 12)(13 51)(12 13)(11 6)
(13 61)(22 14)(21 17)(23 15)(21 13)(23 11)(22 7)(21 9)
(23 7)(22 6)(21 10)(13 12)(22 8)(21 8) ...

Figure 2: Sequence of gazes recorded during performance of an n-term series task

Data of this kind were recorded for 18 think-aloud subjects and 18 silent controls.

3. Task analysis

There has been much research on how subjects solve n-term series problems (see Sternberg, 1980 for an overview). This resulted in a large variety of models for the solution process. These on the one extreme are based upon the idea of subjects building a mental image of the sequential arrangements of terms in the propositions presented to them. When they answer the final question they "read off" from their mental image (DeSoto, London & Handel, 1965; Huttenlocher, 1968). On the other extreme is that position which claims that subjects extract information for each term separately and then answer the final question on the basis of integrating all individual pieces of information (Clark, 1969). More recently, it has been argued that we should not try to determine which of these models is the right one, but rather pay attention to the time course of strategy choice during repeated trials (Johnson-Laird, 1972). Wood, Shotter & Godden (1974) have shown that subjects tend to begin with an imagery strategy and during succesive trials change over to an analytical strategy.

Most research concentrated upon tasks using the left/right dimension, asking subjects questions with respect to the extreme ends of the series. This leaves room for quite a number of strategies. By choosing Ohlsson's way of concluding the experimental tasks (asking for a specific position within the series), the field can be narrowed down to two strategies. These in one case go back to the original (DeSoto, London & Handel, 1965; Huttenlocher, 1968) imagery model: Series Formation, and in the other are a combination of Quinton and Fellows' (1975) and Clark's (1969) model: Elimination. Both strategies at their top level can easily be described as LISP programs.

First, this is the definition of the Series Formation strategy:

```
(DEFUN SF (PROP-LIST QUESTION)

  (READ-OFF (SERIATE PROP-LIST) (CADR QUESTION) (CAR QUESTION)))
```

where PROP-LIST contains the propositions of a task (terms symbolized by "T1", "T2" ...; relations by "R" or "L"), and QUESTION is a list contining a relation and a term. READ-OFF builds a series using SERIATE and returns that element of the series which is connected to the target (CADR QUESTION) through the target relation (CAR QUESTION). SERIATE is defined as:

```
(DEFUN SERIATE (PROP-LIST)

  (DO ((SERIES (PREFERRED-DIRECTION (CAR PROP-LIST)))
       (P-LIST (CDR PROP-LIST)))
  ((NULL P-LIST) SERIES)
  (SETQ SERIES (INTEGRATE SERIES (PREFERRED-DIRECTION (CAR P-LIST))))
  (SETQ P-LIST (CDR P-LIST))))
```

where PREFERRED-DIRECTION returns the two terms of a proposition in the (preferred) left-to-right order and INTEGRATE returns a new series from an old one combined with a new pair of terms in left-to-right order.

Next is the definition for the Elimination strategy:

```
(DEFUN EL (PROP-LIST QUESTION)

   (DO ((ANSWER NIL)
        (ANSWER-TEMP NIL)
        (TARGET-RELATION (CAR QUESTION))
        (TARGET (CADR QUESTION)))
   ((NULL PROP-LIST) ANSWER)
   (COND ((MEMBER TARGET (CAR PROP-LIST))
            (SETQ ANSWER-TEMP (GET-ANSWER TARGET (CAR PROP-LIST)
                               TARGET-RELATION))
             (SETQ ANSWER (COND ((NULL ANSWER-TEMP) ANSWER)
                                 (T ANSWER-TEMP)))))
   (SETQ PROP-LIST (CDR PROP-LIST))))


(DEFUN GET-ANSWER (TARGET PROP RELATION)

   (COND ((NULL PROP) NIL)
         ((EQUAL (CADR PROP) RELATION)
            (COND ((EQUAL (CADDR PROP) TARGET)(RETURN (CAR PROP)))
                  (T NIL)))
         (T (COND ((EQUAL (CADDR (CONVERT PROP)) TARGET)(CADDR PROP))
                  (T NIL)))))


(DEFUN CONVERT (PROP)
    (LIST (CADDR PROP)
          (COND ((EQUAL (CADR PROP) 'R) 'L)
                (T 'R))
          (CAR PROP)))
```

These two strategies have to be extended so that they can cope with an increase in task difficulty: propositions need not be presented in the order of sequential overlap of their terms. As an example, the task in Figure 1 can be presented as shown in Figure 3:

SEVERAL BOYS ARE SITTING ON A BENCH:

KEITH IS LEFT OF CHUCK
MARK IS LEFT OF PHIL
CHUCK IS LEFT OF ROY
MARK IS RIGHT OF ROY

Figure 3: n-term series task without sequential overlap of terms

In order to process such tasks, an additonal function is required for Series Formation which checks whether propositions overlap (function: OVERLAP?). Using this function, there are two alternatives for extending the definition of SERIATE: 1) SERIATE-DISCARD: if a second premise does not overlap, then the old series is discarded and a new series is built; the old proposition will have to processed again later, 2) SERIATE-2ND-SERIES: if there is no overlap, a second series is built which has to integrated into the first one after all propositions have been processed.

```
(DEFUN SERIATE-DISCARD (PROP-LIST)
  (DO ((SERIES (PREFERRED-DIRECTION (CAR PROP-LIST)))
       (P-LIST (CDR PROP-LIST)))
  ((NULL P-LIST) SERIES)
  (COND ((OVERLAP? SERIES (CAR P-LIST))
          (SETQ SERIES (INTEGRATE SERIES
                             (PREFERRED-DIRECTION (CAR P-LIST))))
          (SETQ P-LIST (CDR P-LIST)))
        (T (SETQ SERIES (PREFERRED-DIRECTION (CAR P-LIST)))
           (SETQ P-LIST
                (APPEND (CDR P-LIST)(LIST (CAR PROP-LIST))))))))
```

```
(DEFUN SERIATE-2ND-SERIES (PROP-LIST)

  (DO ((SERIES1 (PREFERRED-DIRECTION (CAR PROP-LIST)))
       (P-LIST (CDR PROP-LIST))
       (SERIES2)(PAIR))
  ((NULL P-LIST)(COND (SERIES2 (INTEGRATE SERIES1 SERIES2))
                      (T SERIES1)))
  (SETQ PAIR (PREFERRED-DIRECTION (CAR P-LIST)))
  (COND ((OVERLAP? PAIR SERIES1)
          (SETQ SERIES1 (INTEGRATE SERIES1 PAIR)))
     (T (SETQ SERIES2 (INTEGRATE SERIES2 PAIR))))
  (SETQ P-LIST (CDR P-LIST))))
```

For Elimination, the situation is more simple, because checking for the occurrence of target terms is already part of the strategy, and overlap therefore does not have any importance.

Before these models can be used any further, they have to be judged with respect to their plausibility. One important qestion is, whether they use storage mechanisms that are compatible with our knowledge of human memory. In the case of Series Formation, an analog representation is postulated which holds up to five items in sequential order that can be rehearsed and recalled over a short period of time. In the case of Elimination, three items have to be kept available for immediate access: TARGET, TARGET- RELATION, and ANSWER; during intermediate steps, one additional temporary variable is required: ANSWER-TEMP. No matter what limitations and mechanisms of short term memory we assume: both these assumptions are extremely plausible.

Another question is related to the complexity of the operations used in the models. This is not critical in the present case: the most difficult operations are those concerning the conversion of relational information, i.e. from "T1 is left of T2" to "T2 is right of T1", and we do consider humans capable of this.

4. Strategy identification

In contrast to earlier approaches (Deffner, 1985) where verbal descriptions of information processing models of task performance served as a basis for strategy identification, the programs Series Formation and Elimination can be used in the present case. Input and output to and from operators in these models can be seen in analogy to attentional proccesses during task performance. But not all attentional processes are observable, since reference to internally represented information need not be related to observable behaviour. Access to items in the external display nevertheless is accompanied by overt behaviour: gazes on these items can be understood to be indicators of such attentional processes, and the overall gaze sequence is a sequence of items attended to.

One note of caution is necessary, though. Gaze direction is not a definite indicator of subjects attending to the item looked at. At least on the level of fixations, perceptual processes determine some eye-movements (c.f. Groner 1978), and also there is the possibility of 'empty stares' where a gaze is not at all directed at the item on the display. For the present purposes, gaze direction nevertheless remains the richest source of data on attentional processes as they are related to visually displayed tasks.

The basic rationale of the present strategy identification is that of matching observed gaze sequences against sequences predicted by models which stand for different strategies. For each model, a program trace can provide this prediction. The degree to which an observed sequence of gazes resembles traces from these models will be used as a basis for quantitative evaluation.

4.1. Derivation of ideal sequences

In order to establish these ideal sequences, both simulation programs were run on all experimental tasks and traced. Tasks 3,4,7, and 8 were of the non-sequential-overlap type, of which SERIATE-DISCARD could solve tasks 3 and 4 and only SERIATE-2ND-SERIES could solve the last two tasks. These more complex versions of SERIATE were used for the task which required them. Only those functions were traced, which can without doubt be considered information processing stages involving visual input. This rules out processes which have to be considered perceptual, and also it rules out any processes involving symbolic processing without immediate or clear reference to the displayed task (c.f. storage/rehearsal mechanisms).

For the SF-program, functions PREFERRED-DIRECTION, OVERLAP?, INTEGRATE, and READ-OFF were traced. Figure 4 shows a trace for the experimental task from figure 3 where terms are symbolized by "T1" through "T5" and the relations left and right by "L" and "R" respectively.

In the case of EL, only GET-ANSWER and CONVERT were traced. MEMBER also is an important function in the program, but it was considered too close to fast perceptual processes to be relevant in the present context of slower information processing stages. Figure 5 presents a trace of EL on task 7.

In a next step, elements in the trace were matched to gaze positions and segmented into stages. Gaze positions within each stage were treated as having equal probability of being looked at, and were consequently joined into sets of possible gaze locations for a given stage. The sequences of sets made up ideal patterns of items attended to under the one or the other strategy. Figure 6 presents an example for task 7:

```
* (SF TASK7 '(R T4))
Entering: PREFERRED-DIRECTION, Argument list: ((T1 L T2))
Exiting: PREFERRED-DIRECTION, Value: (T1 T2)
Entering: PREFERRED-DIRECTION, Argument list: ((T4 L T5))
Exiting: PREFERRED-DIRECTION, Value: (T4 T5)
Entering: OVERLAP?, Argument list: ((T4 T5) (T1 T2))
Exiting: OVERLAP?, Value: NIL
Entering: INTEGRATE, Argument list: (NIL (T4 T5))
Exiting: INTEGRATE, Value: (T4 T5)
Entering: PREFERRED-DIRECTION, Argument list: ((T2 L T3))
Exiting: PREFERRED-DIRECTION, Value: (T2 T3)
Entering: OVERLAP?, Argument list: ((T2 T3) (T1 T2))
Exiting: OVERLAP?, Value: T
Entering: INTEGRATE, Argument list: ((T1 T2) (T2 T3))
Exiting: INTEGRATE, Value: (T1 T2 T3)
Entering: PREFERRED-DIRECTION, Argument list: ((T4 R T3))
Exiting: PREFERRED-DIRECTION, Value: (T3 T4)
Entering: OVERLAP?, Argument list: ((T3 T4) (T1 T2 T3))
Exiting: OVERLAP?, Value: T
Entering: INTEGRATE, Argument list: ((T1 T2 T3) (T3 T4))
Exiting: INTEGRATE, Value: (T1 T2 T3 T4)
Entering: INTEGRATE, Argument list: ((T1 T2 T3 T4) (T4 T5))
Exiting: INTEGRATE, Value: (T1 T2 T3 T4 T5)
Entering: READ-OFF, Argument list: ((T1 T2 T3 T4 T5) T4 R)
Exiting: READ-OFF, Value: T5

T5
```

Figure 4: Trace of SF on task 7

```
* (EL TASK7 '(R T4))
Entering: GET-ANSWER, Argument list: (T4 (T4 L T5) R)
 Entering: CONVERT, Argument list: ((T4 L T5))
 Exiting: CONVERT, Value: (T5 R T4)
Exiting: GET-ANSWER, Value: T5
Entering: GET-ANSWER, Argument list: (T4 (T4 R T3) R)
Exiting: GET-ANSWER, Value: NIL

T5
```

Figure 5: Trace of EL on task 7

SF: ((11 12 13)(21 22 23)(31 32 33)(41 42 43)(52 53 23))
EL: ((52 53)(21 22 52 53 23)(42 43 52 53 23))

Figure 6: ideal patterns for SF and EL on task 7

There was one more complication, however: with the Elimination method, there is no need to use only one set order in which the propositions are processed. Though it does not seem reasonable to assume that the order is random, it must be conceded that there are two plausible orders. In one, subjects work their way down the list of three or four propositions from the top to bottom, whereas in the other case, they start from the bottom line and work upwards. For this reason, there have to be two ideal sequences for EL, thus a third pattern has to be added to Figure 6: EL-2: ((52 53)(42 43 52 53)(21 22 52 53 23)).

4.2. Matching gaze sequences to program traces

Matching was straightforward: Starting with the first list in the pattern, its elements were checked against successive elements of the observed gaze sequence. A match started with the first element from that list and was continued until more than one successive element in the gaze sequence was extraneous to the list from the pattern, or the duration of an individual extraneous gaze was longer than the average gaze duration in the total gaze sequence. When a match was discontinued, the next list from the pattern was used; if the pattern was exhausted, matching started from the beginning of the pattern. Because of this strict sequential order in which lists from the pattern were matched, only such backup in task performance could be identified which would start from the very beginning. All attempts to allow for partial backup resulted in substantial loss of clearness of strategy identification, and were not included in the final version of the algorithm.

This algorithm was used three times for each gaze sequence: once for the Series Formation strategy and twice for the Elimination strategy (using the forward and the backward pattern). The numerical information used for further data analyses consisted of the percentage of total gaze sequence which could be matched to each pattern. For Elimination, only the higher of the two percentages was used, so that differences between backward or forward task performance could be ignored in the overall comparison of Elimination to Series Formation. Another extension was that these analyses were performed for both gaze duration and gaze frequency as a basis of percentages. The reason for using both these measures was that no plausible argument could be found to favor either one or the other on theoretical grounds. Table 1 presents an example of the resulting output.

Table 1: Sample output of percent matches

Subj.	%SF/dur	%EL/dur	%SF/freq	%EL/freq
S-EL	0.00	59.38	0.00	46.50
10	86.89	25.40	86.17	14.17
11	0.00	68.75	0.00	38.02
13	95.45	40.91	99.08	48.31
14	38.03	8.22	44.18	9.35
15	29.41	17.07	28.64	14.07
17	60.87	10.20	59.68	9.64
18	46.51	10.20	36.63	5.55

As can be seen from Table 1, using duration or frequency as a basis of percentages did not result in great differences, the two measures are highly correlated. Also, the percentages for EL tended to be smaller. This argues for less appropriateness of the patterns and/or pattern matching used for EL. These figures nevertheless could be used well, the only consequence was that they should not be treated as variables on the same dimensions.

Two methods were used to extract information from the four variables per person and task: principal components analysis and cluster analysis. The former resulted in very clear factor structures with one factor explaining a large proportion of the variance. It was identified as a bipolar SF - EL factor and scores on this factor were used for analyses involving continously scaled numercial information.

Cluster analysis was used as a basis of a binary categorization of individual task performances with respect to their predominant strategy. Using the k-means method, it was possible to obtain clear two-cluster solutions for all eight tasks. In all eight cases, these could easily be identified as an Elimination cluster and a Series Formation cluster. Cluster membership was then used for the categorization of task performance.

5. Validity

The validity of these measures was checked in two ways. Firstly, the analysis was performed on additional data recorded from subjects who had received prior training in one of the strategies. Out of a group of six, there were only two subjects who in their subsequent judgement were positive of having used nothing but the trained strategy - one subject trained to use Elimination (called S-EL) and one trained to use Series Formation (S-SF). Table 2 presents the percentage of successful matches for these two sets of data.

Table 2: percent matches for two trained subjects

	Subj.	%SF/dur	%EL/dur	%SF/freq	%EL/freq
Task1	S-EL	0.00	59.38	0.00	46.50
	S-SF	75.00	29.17	83.71	11.62
Task2	S-EL	0.00	47.06	0.00	58.24
	S-SF	80.95	4.76	85.97	2.20
Task3	S-EL	0.00	100.00	0.00	100.00
	S-SF	59.46	27.03	65.11	14.74
Task4	S-EL	0.00	91.67	0.00	98.02
	S-SF	67.86	7.14	83.65	2.71
Task5	S-EL	0.00	60.00	0.00	76.95
	S-SF	42.86	6.98	49.47	3.35
Task6	S-EL	0.00	63.64	0.00	54.96
	S-SF	62.07	6.90	75.61	1.39
Task7	S-EL	0.00	35.29	0.00	30.32
	S-SF	76.20	7.94	69.89	9.16
Task8	S-EL	0.00	25.00	0.00	23.63
	S-SF	59.25	9.09	52.02	6.44

As can be seen from Table 2, there is very good separation between S-SF and S-EL.

The other line of approach was based on an idea used by Wood et al. (1974). These authors surprised subjects with a repeted presentation of a task where in the repetition they did not ask for the position of one specific person, but required subjects to give the total arrangement instead. The difference in solution time between the first and the second solution of the task was used to estimate whether subjects had been using the Series Formation strategy - long times on the repetition standing for prior use of the Elimination strategy. In the present case, solution time on the repetion of task 8 (expressed as a factor of solution time for task 8) was compared for subjects whose solution attempt had been classified as SF or EL on the basis of cluster analyses. Table 3 shows means scores:

Table 3: Means scores for relative solution time on task 9

	$\bar{x}$	s	
SF (N = 18)	2.72	3.05	t = 3.908 p < .01
EL (N = 17)	7.04	4.65	

There is a clear difference in the expected direction: subjects whose solution of task 8 was categorized as Series Formation required significantly less time when asked for the total arrangement.

Thus, both approaches to testing the validity of strategy identification gave very clear and reassuring results.

6. Conclusion

Instead of arguing for the appropriateness of the assumptions underlying the approach to the identification of strategies presented here, I shall present a brief glimpse at the results.

The comparison of mean factor scores (Factor SF-EL) revealed significant differences in the case of task 3 and 4: mean scores were higher in the silent group (indicating more use of the Elimination strategy). This is borne out by the categorizations on the basis of on the basis of cluster analyses, which lend themselves more readily for graphical presentation: Figure 7 shows the frequency of Elimination in the two groups.

These differences can be interpreted as follows: There is no big difference between the two gropus with respect to strategy use. What is different, is the speed at which they discover the Elimination strategy: Thinking-aloud subjects are slower.

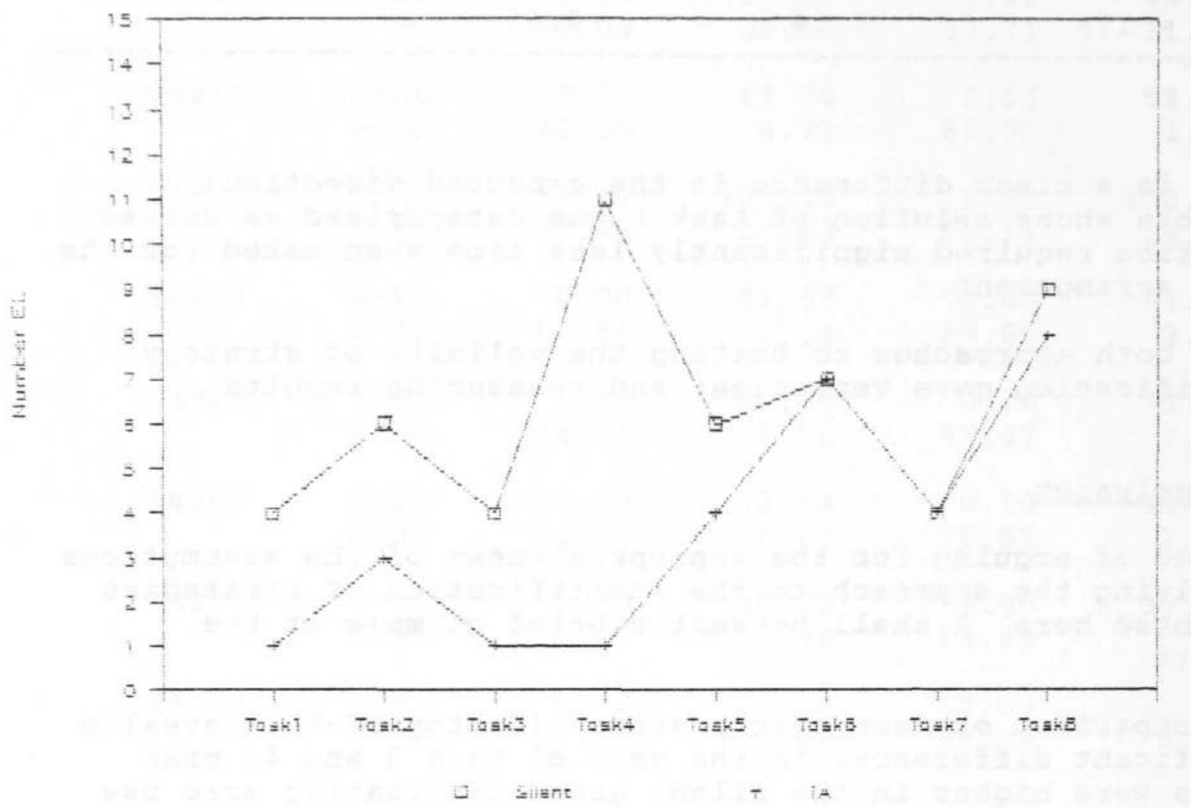

Figure 7: Use of the Elimination strategy over 8 tasks

References

Clark, H.H. (1969) Linguistic Processes in deductive reasoning. Psychological Review, 76, 387-404.

Deffner, G. (1984) Lautes Denken - Untersuchung zur Qualität eines Datenerhebungsverfahrens. Frankfurt: Lang.

Deffner, G. (1985) Identification of solution strategies on the basis of eye-movement data. In R. Groner, G.W. McConkie & C. Menz (Eds.), Eye movements and human information processing. Amsterdam: North Holland.

DeSoto, C.B., London, M. & Handel, S. (1965) Social reasoning and spatial paralogic. Journal of Personality and Social Psychology, 2, 513-521.

Groner, R. (1978) Hypothesen im Denkprozess. Bern: Huber.

Huttenlocher, J. (1968) Constructing spatial images: A strategy in reasoning. Psychological Review, 7, 351-357.

Johnson-Laird, P.N. (1972) The tree-term series problem. Cognition, 1, 57-82.

Just, M.A. & Carpenter, P.A. (1976) Eye fixation and cognitive processes. Cognitive Psychology, 8, 441-480.

Kliegl, R. (1981) Automated and interactive analysis of eye fixation data in reading. Research Methods & Instrumentation, 13, 115-120.

Kliegl, R. & Ohlsson, R.K. (1981) Reduction and calibration of eye monitor data. Behavior Research Methods & Instrumentation, 13, 107-111.

Ohlsson, S. (1982) Competence and strategy in reasoning with common spatial concepts (Working Papers from the Cognitive Seminar, No 6) Stockholm: Department of Psychology.

Quinton, G. & Fellows, B.J. (1975) 'Perceptual' strategies in the solving of three-term series problems. British Journal of Psychology, 66, 69-78.

Sternberg, R.J. (1980) Representation and process in linear syllogistic reasoning. Journal of Experimental Psychology: General, 109, 119-159.

Wood, D. Shotter, J. & Godden, D. (1974) An investigation of the relationships between problem solving strategies, representation and memory. Quarterly Journal of Experimental Psychology, 26, 252-257.