The effects of irrelevant relations on the processing of sequential patterns*

HARRY M. HERSH

Johns Hopkins University, Baltimore, Maryland 21218

Letter series completion problems varying along location of irrelevant relations (beginning, end, none) were presented to 24 Ss. A repeated-measures analysis of solution times and errors revealed that irrelevant relations at the beginning of series produced the longest latencies and the most errors. Using a pattern induction/sequence production model of serial processing, an error analysis showed production error rate remained constant across treatments, while pattern induction errors varied with confusion location. The results indicate that the difficulty of sequential processing varies with irrelevant relations only during the pattern induction phase. The findings also support a left-right generate-and-test model of pattern induction for the task.

Many theories have been advanced concerning the processing of sequential patterns (e.g., Gregg, 1967; Klahr & Wallace, 1970; Restle, 1970; Simon & Kotovsky, 1963; Simon & Sumner, 1968; Vitz & Todd, 1969; Williams, 1969). Simon (1972) has argued that each of the theories is in agreement with the hypothesis that Ss induce a pattern description from the sequence and use this description in the production of new sequences. In addition, he argued that pattern length was a reasonable predictor of processing difficulty, independent of task and theoretical coding scheme.

One popular context for the study of sequential patterns has been the Thurstone Letter Series Completion Test (Klahr & Wallace, 1970; Kotovsky & Simon, 1973; Simon & Kotovsky, 1963; Thurstone & Thurstone, 1941). In the task Ss are shown a series of letters (e.g., A A B B C C D —) and are asked to continue the series for the specified number of letters. It is assumed that a S solves the problem by first inducing a pattern description of the series. Simon and Kotovsky (1963) contend that Ss arrive at this pattern description by first discovering the periodicity, or cycle length, of the sequence. The period, in turn, is found by looking for a relation which either repeats or is interrupted at regular intervals. Once the period is discovered, the pattern is completed by detecting and organizing the remaining relations which hold between letters within a period and/or corresponding letters across periods.

The Klahr and Wallace (1970) model employs a matching procedure for arriving at a comparable pattern description. The first item in the series is tested against the second for the presence of a legitimate relation (e.g., a repeated letter or a contiguous relation in the alphabet). If a relation is found, the model then determines whether it holds for the entire series. If no relation is found, then the first two items in the series

become the template to be matched. The procedure continues until a template is found which holds across the series. Thus, implicitly, the Klahr and Wallace model also seeks to discover the periodicity of the sequence.

Each model assumes a left-right search for relations which are then tested across the series and finally combined to form the pattern description. Both also agree that, once the description is discovered and organized, it is then used to extrapolate the series.

Although such factors as pattern length, pattern complexity, and short-term memory requirements have all been investigated and found to contribute to the overall difficulty of the task, one variable that has been mentioned (Simon & Kotovsky, 1963, p. 544; Kotovsky & Simon, 1973, p. 401) but not investigated systematically is the influence of irrelevant relations. For example, Simon and Kotovsky's model would judge the two series M M M N M O and A M A N A O to be of equal difficulty, since each can be generated by the same pattern description (a repeated letter alternating with a progression of the alphabet). However, even to the casual observer, the second series appears much easier to continue than the first.

The present research was concerned with an analysis of the difficulty of letter series completion problems as a function of irrelevant relations (and the resulting confusion). Assuming that Ss typically process the sequences from left to right (see also Neisser, 1967, p. 37), irrelevant relations occurring toward the beginning of a series will increase the number of hypotheses to be tested. The additional relations should increase solution times, since more hypotheses typically will have been tested before a relation is found that holds across the series. With more relations to test, the possibility of inducing an incorrect pattern description should also increase since (1) the probability of finding a relation that holds across the series will decrease and (2) the possibility of accepting an incorrect hypothesis will increase if such hypotheses are not tested across the complete series.

*This research was supported by PHS Grant 5 T01 GM02148 to Johns Hopkins University. The author wishes to express appreciation to Bert F. Green, Jr. for his helpful suggestions in regard to this research project.

Table 1
Rules Used to Generate the Pattern Descriptions

Relation	Symbol	Example
Same	S	D D
Next	N1	R S
Next of Next	N2	A C
Backward Next	B	V U

Irrelevant relations occurring toward the end of a sequence should also increase problem difficulty. While there will be fewer hypotheses to be tested than when the irrelevant relations (and confusion) occur at the beginning, these relations will make accepting an incorrect hypothesis easier, due to the possibility that such hypotheses might be supported by spurious but similar relations occurring toward the end. The increased confusion resulting from the irrelevant relations at the end of a series should also result in increased solution times.

Logically, problem difficulty can be encountered during either the pattern induction phase or the series extrapolation (production) phase. Since it is argued that the irrelevant relations affect only the induction phase (i.e., hypothesis generating and testing operations), then, given that a correct pattern has been induced, errors occurring during the production phase should not be affected by the irrelevant relations.

METHOD

Stimuli

Four elementary relations were used to generate the test sequences: same (S), next (N1), next of next (N2), and backward next (B) (see Table 1). The relations were randomly grouped into 30 pattern descriptions, 15 of Length 3 and 15 of Length 4. Thus, a typical pattern of Length 3 might be S S N1.

Periods (i.e., pattern description lengths) of three and four were chosen to avoid experimental set, while also allowing the investigation of the effect of period length on sequence difficulty. Letters in the initial period of each sequence were also chosen randomly. Thus, the description S S N1 might produce E F I E F J E F K _ _ _ or A X Q A X R A X S _ _ _.

Each pattern description was in fact used to generate three related letter series. Two variations on the original series were obtained by changing one of the letters (or one related pair of letters) in the initial period of each series. The letters in the modified sequences were changed such that one sequence contained irrelevant relations at the beginning, while another now had possible irrelevant relations at the end. For example, E F I could become E F E or E F C, producing E F E E F F E F G and E F C E F D E F E using the pattern description S S N1. In both modifications, one irrelevant relation is the alternating E, the first occurring toward the beginning of the series, the second toward the end. The resulting 90 sequences (3 each of 30 patterns) are shown in Table 2. Series 1-45 are of Period 3, 46-90 of Period 4.

Subjects

Twenty-four students at Johns Hopkins University, 12 male and 12 female, served as Ss and were paid for participating.

Table 2
Ninety Letter Series

1. MM00OPQQQS	31. SSRRRRQQRPP	61. IJIIJHIJGI
2. MMROOSQOTS	32. SSTRRTQQTTP	62. IJIIJGIJFI
3. MMQOORQQSS	33. SSORROQOPP	63. IJLIJKIJJ
4. DEEDDDCCDB	34. JKLLKLNKLPKL	64. EFEFEFFGEFGH
5. FEEFDDFCCFB	35. MKLOKLOKLSKL	65. EFFGEFGHEFHI
6. BEEBDDCCBB	36. DKLFKLHLKJL	66. EFBCEFCDEFDE
7. STUTUTUVSVW	37. WVWVWVWUWVTW	67. MMLLLMMMKNNJ
8. STRTUQVVPVW	38. WVUWVTWVSWVRW	68. MMNNLLOOKKPPJ
9. STYTUXUVWVW	39. WVZWVYVWXVWVW	69. MMILLJJKKKKJ
10. SSRRSRQSRPSR	40. FEEFFEFGEFHEF	70. KJJKKLJMKNJ
11. QSRPSROSNSR	41. FGEFHEFIEFJEF	71. KKJLKMJNKOJ
12. VSRUSRTSRSSR	42. FBFECEFCDEFEEF	72. KFJGKHJIKIJ
13. EGEGGEIGEK	43. CDEDEDEFCEFG	73. AABBBDDCCF
14. FGEHGEJGEL	44. CDKDEJEFIFGH	74. EEBBFFDDGGF
15. AGECEGEGEG	45. CDIDEHEFGFGF	75. DDBBEEDDFFF
16. EFFFHGHGHI	46. LLMMMLLNK	76. NNNNMNNLNL
17. EFAFGCGHEHI	47. LLKMMJJNNI	77. NONOMOLOL
18. EFCFGEHGH	48. LLPPMMOONNN	78. NLNLMLMLLL
19. RPQQRSTUOVW	49. DAADDACDDAE	79. DEDEEFFGFGHI
20. PPQQRSTUMVW	50. DAAFDACFDAE	80. DEFGEFHIFGJK
21. ZPQYRSXTUVW	51. DAAEDACEDAE	81. DECDEFEEFFGGH
22. MNMNNNOONPN	52. XXXXXWVXXV	82. FDDFFFFFHFF
23. MNPNNRONTN	53. XXWXXVXXUU	83. FEEFFGGFFIF
24. MNJNNLONNPN	54. XXZZXXYXXXX	84. FBBFFDDFFFFF
25. TTUUSTVRSWQ	55. LKLLKKLJKJL	85. HIJHIKJHJKLH
26. VTUWSTXRSYQ	56. LMLKMKLJMJL	86. HIJGJGJKJL
27. NTUOSTPRSQQ	57. LJLLKJLJL	87. HIJKJJKJL
28. FGHGHHHIIJ	58. DEDCECCDEB	88. RSSTSTSTTUST
29. FGEHHEHIEJ	59. DEFCECED	89. TUSTVSTVWST
30. FGJGHJHIJ	60. DEECDEDCDEC	90. QRSTRSSTSTST

Procedure

The sequences were randomly placed in three groups of 30. The series in each group were randomly selected, subject to the restrictions that (1) each group contained exactly one variation of each pattern description and (2) each group contained 10 sequences of each variation (no irrelevant relations, irrelevant relations toward the beginning, irrelevant relations toward the end), five of Period 3 and five of Period 4. Each S was presented with one of the three groups, such that every group was seen by eight Ss.

The experiment was computer controlled using a DEC PDP-11/20 and a Techtronics CRT terminal. The computer determined randomly which group of sequences would be presented to each S. It was also used to display instructions to the S.

The S was told that 30 letter series would be presented, one at a time. Three blanks would follow each series. The S was instructed to continue the series by typing the next three letters in the series. Although finding correct answers was emphasized, Ss were made aware that their times to solution were being recorded by the computer.

Eight practice problems of increasing difficulty followed the instructions. The 30 test problems were then presented, one at a time. Ss could work on a problem as long as they desired. A carriage return, indicating the S had no solution, stopped the clock and caused the answer to be recorded as a series of three question marks.

RESULTS

The data were analyzed in terms of eight random sets of three Ss each (one from each stimulus group), thus allowing a two-factor repeated-measures design to be applied to the sets. Mean latencies over S sets and over sequences within conditions are shown in Fig. 1. Mean errors over sets of Ss and summed across sequences are shown in Fig. 2. The product-moment correlation between errors and solution times over the 90 series was moderate [$r(88) = .566, p < .001$]. Thus, both latencies and errors appear to be indicants of the common underlying process.

Fig. 1. Mean latencies as a function of period length and location of irrelevant relations.

Fig. 2. Mean errors as a function of period length and location of irrelevant relations.

An analysis of the latency data (after a log transformation) indicated a significant difference due to the location of the irrelevant relations [$F(2,14) = 34.60, p < .001$]. There was no difference in solution times as a function of period length. A post hoc analysis (Scheffé's method) showed a significant difference in latencies between each pair of relation loci ($p < .05, \text{time}_{\text{begin}} > \text{time}_{\text{end}} > \text{time}_{\text{none}}$).

Error rate also varied with the location of irrelevant relations [$F(2,14) = 15.04, p < .001$]. For the error data, there was a significant interaction of Location of the Irrelevant Relation by Period Length [$F(2,14) = 4.33, p < .05$]. The latency data showed a similar trend; however, the result was nonsignificant [$F(2,14) = 3.58, p > .05$].

Two objective rules were used to determine which errors would be classified as occurring during the production phase: (1) If the extrapolation was out of phase in reference to the period, it indicated that, although the S had induced a correct pattern description, he had trouble keeping his place in the series during the extrapolation process. (2) If two of the three letters were correct but the third was one letter out of phase on the alphabet, it indicated that the error was a function of a problem in reproducing the alphabet and not in inducing the pattern. All other errors were considered to be conceptual.

Conceptual and production errors over Ss are shown in Fig. 3. The production error rate was constant across treatments. The conceptual error rate varied significantly as a function of location of irrelevant relation [$F(2,14) = 20.38, p < .001$]. A post hoc analysis (Scheffé's method) revealed significant differences between the conceptual error means for each treatment ($p < .05, \text{error}_{\text{begin}} > \text{error}_{\text{end}} > \text{error}_{\text{none}}$).

DISCUSSION

The finding that irrelevant relations at the beginning of a sequence increases the solution time supports the contention that Ss process from left to right, testing discovered relations (relevant or otherwise) until one is found that is consistent across the series. When many of the discovered relations are found not to hold (or hold only across part of the series), then solution times increase markedly. An additional cause of the increase in latencies is that Ss may be retesting false hypotheses. Huesmann and Cheng (1973) have shown that in a mathematical induction task Ss often retest a hypothesis that has been rejected if the hypothesis was partially supported during a previous test (on the same problem).

The variation in error rate over location of the irrelevant relation indicates that, when an irrelevant relation occurs at the beginning of a series, Ss are more apt to induce incorrect or incomplete pattern descriptions. The incorrect descriptions can have three sources: (1) Each relation was correctly discovered and tested, but the S had problems in organizing the relation into a coherent rule. (2) The S accepted an incorrect or partially correct relation due to a failure to test the relation along the complete sequence. (3) The S accepted a partially correct relation because more complete patterns could not be discovered. The acceptance of partially consistent relations can be clearly seen in many of the incorrect sequence extrapolations.

Production errors appeared to be independent of location of the irrelevant relation, implying that, once a pattern description had been correctly induced and organized, the error rate was a function only of pattern complexity.

The interaction found between period length and location of the irrelevant relation appears to be an experimental artifact. The mean length of the letter sequences was constant across periodicity, so series length could not be used as an index of period length. However, Ss had fewer periods to analyze within a given series when the periodicity was four. The fewer periods per series resulted in reduced errors with no irrelevant relations at the beginning of the series. With irrelevant relations at the end and the mean number of periods less than three (for a periodicity of four), Ss found it more difficult to verify the relations and, therefore, produced more errors. The confounding between periodicity and sequence length also resulted in no significant differences in solution time or error rate between periods of Length 3 and Length 4.

Simon (1972) has shown that the many sequential processing theories are actually the same. Given the diversification of tasks to which these theories address themselves (e.g., music, binary choices, letter series), it seems reasonable to assume that the occurrence of

Fig. 3. Conceptual and production errors as a function of irrelevant relations.

irrelevant relations, or similar confusions and ambiguities, in sequential patterns greatly increases the difficulty of correctly processing such patterns. The influence is especially noticeable when the irrelevant relations occur toward the beginning of the pattern. Moreover, the difficulty of processing in the presence of irrelevant relations occurs only during the induction phase: Once the pattern or rule has been induced, the production is unaffected by the relations.

REFERENCES

- Gregg, L. W. Internal representation of sequential concepts. In B. Kleinmuntz (Ed.), *Concepts and the structure of memory*. New York: Wiley, 1967.
- Huesmann, L. R., & Cheng, C. A theory for the induction of mathematical functions. *Psychological Review*, 1973, 80, 126-138.
- Klahr, D., & Wallace, J. G. The development of serial completion strategies: An information processing analysis. *British Journal of Psychology*, 1970, 61, 243-257.
- Kotovsky, K., & Simon, H. A. Empirical tests of a theory of human acquisition of concepts for sequential patterns. *Cognitive Psychology*, 1973, 4, 399-424.
- Simon, H. A. Complexity and the representation of patterned sequences of symbols. *Psychological Review*, 1972, 79, 369-382.
- Simon, H. A., & Kotovsky, K. Human acquisition of concepts for sequential patterns. *Psychological Review*, 1963, 70, 534-546.
- Simon, H. A., & Sumner, R. K. Pattern in music. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley, 1968.
- Thurstone, L. L., & Thurstone, T. G. *Factorial studies of intelligence*. Chicago: University of Chicago Press, 1941.
- Vitz, P. C., & Todd, R. C. A coded element model of the perceptual processing of sequential stimuli. *Psychological Review*, 1969, 76, 433-449.
- Williams, D. S. Computer program organization induced from problem examples. Unpublished doctoral dissertation, Carnegie-Mellon University, 1969.

(Received for publication February 7, 1974;
revision accepted April 27, 1974.)