A COMPUTATIONAL MODEL OF ATTENTIONAL REQUIREMENTS IN SEQUENCE LEARNING

Peggy J. Jennings and Steven W. Keele
Department of Psychology
University of Oregon

ABSTRACT

This paper presents a computational model of attentional requirements in sequence learning. The structure of a keypressing sequence affects subjects' abilities to learn the sequence in a dual task paradigm (Cohen, Ivry, & Keele, 1990). Sequences containing unique associations among successive positions (i.e., 1-5-4-2-3) are learned during distraction. Sequences containing repeated positions with ambiguous associations (i.e., 3-1-2-1-3-2) are not learned during distraction. Cohen, et al. proposed two fundamental operations in sequence learning. An associative mechanism mediates learning of the unique patterns (1-5-4-2-3). These associations do not require attention to be learned. Such an associative mechanism is poorly suited for learning the sequence with repeated elements and ambiguous associations. These sequences must be parsed and organized in a hierarchical manner. This hierarchical organization requires attention.

The simulations reported in this paper were run on an associative model of sequence learning developed by Jordan (1986). Sequences of differing structures were presented to the model under two conditions -- unparsed, and parsed into subsequences. The simulations modeled closely the keypressing task used by Cohen, Ivry and Keele (1990). The simulations (1) replicate the empirical findings, and (2) suggest that imposing hierarchical organization on sequences with ambiguous associations significantly improves the model's ability to learn those sequences. Implications for the analysis of fundamental computations underlying a system of skilled movement are discussed.

This paper examines whether a connectionist model of sequence learning developed by Jordan (1986, 1990) provides insight into the attentional requirements of sequence learning as investigated by Cohen, Ivry, and Keele (1990; see also Keele, Cohen, & Ivry, 1990). The first section describes the central phenomena explored by Cohen, et al. The second section describes aspects of Jordan's model and how it relates to the empirically discovered phenomena.

BEHAVIORAL CHARACTERISTICS OF SEQUENCE LEARNING

Cohen, et al. adopted a paradigm established by Nissen and Bullemer (1987) to investigate the learning of sequential representations. Successive presentations of a visual stimulus appeared at one of 3, 4, or 5 locations on a screen. Subjects responded to each stimulus by pressing a key corresponding to the stimulus location. Reaction times were recorded. Unknown to the subjects, a large proportion of the successive signals appeared in a particular repeating sequence of locations. One type of sequence, called Unique, involves a sequence of 5 unique signal positions, an example of which is 1-5-4-2-3. The numbers refer to signal positions. After the last position in a cycle, the sequence repeated with no detectable break. Sequences were presented to subjects in 8 blocks of 20 cycles through the sequence. A second type of sequence, called Ambiguous, involves only 3 signal positions. Each position is repeated within the sequence, but each occurrence is followed by a different successor. An example of an Ambiguous sequence is 1-3-2-3-1-2.

Cohen, et al. found that, with practice, subjects learn both of these types of sequences in the absence of a distraction task. Sequence acquisition was demonstrated by steady improvement in reaction time over 8 blocks of training, with a significant increase in reaction time on blocks of trials presented after training in which the signals occurred at random rather than in the structured sequence. When a secondary task was performed simultaneously, diverting attention from the primary reaction time task, Unique sequences were learned but Ambiguous ones were not. That is, performance to signals occurring in the structured Ambiguous sequence never became faster than performance to randomly occurring signals.

Something about learning the Ambiguous sequence appears to require attention. Cohen, et al. also observed that if an Ambiguous sequence is altered by replacing one of the repeated events with an event that occurs only once in the cycle, such as 1-4-2-3-1-2 (the underline shows the unique event), subjects learn this sequence at a rate similar to that for the Unique sequences. These sequences are called Hybrid because they involve a mixture of repeated and unique events.

Cohen, et al. offered the following explanation of the effects of sequence structure on learning. The Unique sequences can be acquired by forming associations between adjacent events. Such learning appears to require little or no attention to the relationships among items. Associational learning is not well suited for learning Ambiguous sequences, however, because on different occasions a particular event is ambiguously followed by a different event. Ambiguous sequences can, however, be learned by a mechanism that divides the sequence into parts. This process of hierarchic organization seems to require attention. Why the Hybrid sequences are learned more like Unique than Ambiguous sequences is unclear.

A COMPUTATIONAL MODEL OF SEQUENCE LEARNING

The goal of the present research is to gain insight into (1) how the structure of a sequence interacts with component operations of a sequence learning system; and (2) how parsing and hierarchic representation may be implemented in a neural network. These goals were approached using Jordan's (1986, 1990) recurrent network model. His model, as we implemented it, is illustrated in Figure 1. An input layer has units of two types: plan units and context units. These input units are connected to a layer of hidden units, and those hidden units feed into output units that we call prediction units. Activation of these latter units represents a prediction, or priming, of the upcoming event. The prediction units can be viewed variously as representing a prediction of the upcoming stimulus or, because the stimulus determines which response to make, the upcoming response. Stimulus information constitutes the target output that is used to calculate error in the prediction units and forms the basis for

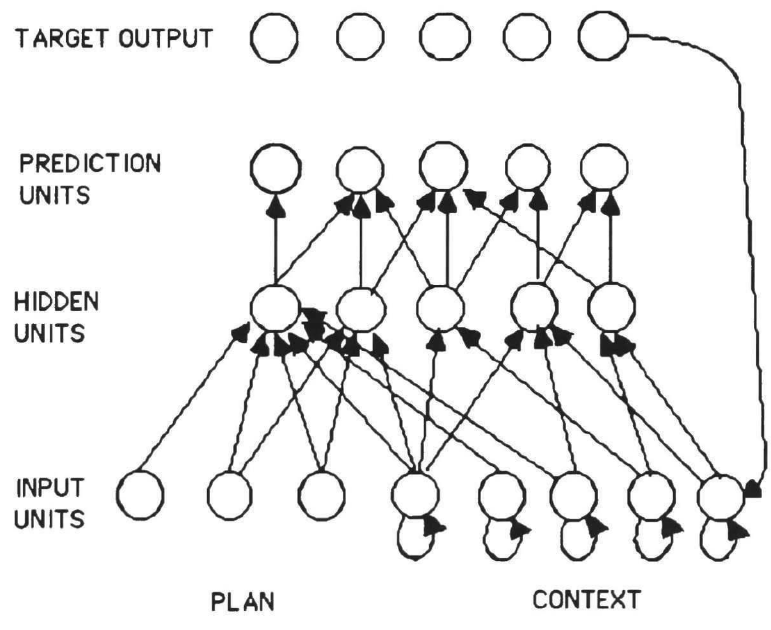

Figure 1. Jordan (1986) recurrent network model. (Not all connections are shown.)

learning using backpropagation. Changes in connection weights are made over the course of learning based on the amount of prediction error on each trial.

During training, information about the target response feeds back to the context units with a fixed connection weight of 1. Output representation is local rather than distributed, and each target output unit is connected to one and only one context unit. Each context unit also feeds a proportion of its previous activation back on itself. The recurrent connections produce context unit activations that retain a history of recent events. In an Ambiguous sequence such as 1-3-2-3-1-2, the context following stimulus 3 is somewhat different each time the 3 occurs because in each case it is preceded by different events. The influence of those events, while diminishing over successive stimuli, is partially retained by their representation in the context units.

Sequence learning is achieved as follows: at the beginning of a sequence a pattern of activation appears on the plan units. In the general case, this pattern persists in unmodified form throughout the training session. At the beginning of a block of training, the context units are all set to zero. The combination of plan and initial context values feed through the hidden layer to produce a prediction. A "stimulus", or target response, provides target output information. Prediction error is calculated, and weight changes occur via backpropagation. The "stimulus" information also feeds back to the context units, and the context units reverberate a portion of their previous values. On subsequent iterations, the updated context values in combination with the plan values constitute the input patterns associated with consecutive target responses.

In these simulations, the model was trained on two distinct types of sequence organization. One condition involved presenting 20 cycles through each sequence type with no higher-level organization. In another condition, the input patterns were parsed into distinct subparts. Two implementations of parsing were examined. One implementation involved resetting the context units to zero at the beginning of each cycle through the sequence. This implements a form of parsing whereby the boundaries of a sequence are marked. The second implementation of parsing involved assigning different plan values to each subpart of a sequence. When a subsequence is finished and a new one is to start, the activation pattern on the plan units changes accordingly. In each implementation, parsing is made explicit in the input patterns.

The intent of these simulations is to determine how parsing influences the learning of the three sequence types: Unique, Hybrid, and Ambiguous. Parsing, in this conception, corresponds in the empirical case to an attention-dependent organizational process that mediates learning of Ambiguous sequences.

Simulation 1 -- Associative Processes in Sequence Learning

The goal of Simulation 1 was to replicate the behavioral characteristics of sequence learning under divided attention as reported by Cohen, Ivry, and Keele (1990). Under dual-task conditions, performance of Unique sequences (e.g., 1-5-4-2-3) and Hybrid sequences (e.g., 1-4-2-3-1-2) steadily improves with training, while performance of Ambiguous sequences (e.g., 1-3-2-3-1-2) remains no better than performance to randomly presented stimuli.

Training Set

Input to the model was composed of 2 elements. (1) a plan value; and (2) a representation of recent target keypress history. Plan values remained constant throughout the training session for each sequence. Target keypress values were represented by a vector of binary values across 5 units. For example, an output pattern of 00100 represents a press of the middle key. Context values represented the sum of the desired output at time t and a proportion of the context value at time t-1. Training sets consisted of 20 cycles through a sequence, consistent with a single block of training in the empirical task. Each input pattern is associated with a desired output value corresponding to a keypress. At the beginning of a block of training, context units were set to zero. Performance measures consisted of (1) total sum of squares of prediction error summed over individual trials and recorded after each block of training; and (2) number of blocks required to learn the sequence to specified accuracy criterion (less than 0.04 sum of squares prediction error over 20 cycles). The model was trained on three exemplars of each sequence type: Unique (1-5-2-4-3; 1-4-5-3-2; 1-3-4-2-5), Ambiguous (1-2-3-1-3-2; 1-2-3-2-1-3; 1-3-2-3-1-2), and a Hybrid of unique and ambiguous associations (1-2-3-1-3-4; 1-2-3-2-4-3; 1-4-2-3-1-2). Six separate training sessions per sequence type were run; two sessions per sequence. Mean values from sequence type groups were analyzed.

Results of Simulation 1

An analysis of variance of number of blocks required to learn the sequence to criterion showed a significant effect of sequence type ($F(2,15) = 19.742$, $p <= 0.001$). Multiple comparisons (Tukey, $p < .05$) reveal that the model requires fewer blocks of training to learn the Unique sequences (mean blocks = 62.2) and the Hybrid sequences (mean blocks = 68.0) than to learn the Ambiguous sequences (mean blocks = 106.0). The Unique and Hybrid conditions are not significantly different.

Figure 2 displays the rate of change in prediction error over the first 10 blocks of training. While learning the Unique sequences, the reduction in prediction error is rapid and steadily decreasing with training, approaching zero. While learning the Amibiguous sequences, however, prediction error levels off at a relatively high value after an initial brief and rapid decline. These error measures may be compared with improvement in reaction time reported in the empirical case, resulting from the subject's increasing ability to correctly anticipate events. In the Ambiguous condition, prediction error as well as empirical reaction times remain relatively high as anticipation of response is not much better than chance.

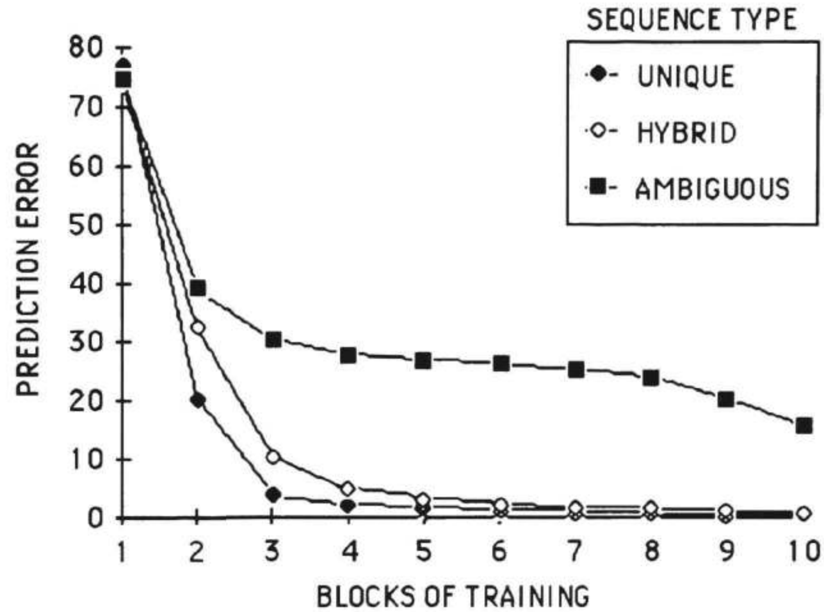

Figure 2. Prediction error as a function of training, by sequence type.

Discussion

Learning in the simulation is denoted by reductions in error between prediction values and target response values while learning in humans is documented by reductions in reaction time. Human reaction time is improved by priming or expectancy. Model error as well as empirical reaction times, because block means are reported, reflect accuracy of prediction over the course of 20 cycles through a sequence.

In this simulation, both Unique and Hybrid sequences were learned at essentially the same rate. These results agree qualitatively with the results of human learning in which learning of Unique and Hybrid sequences also occurs at similar rates when a distraction task is used. In contrast to the equivalence of Unique and Hybrid sequence learning rates, the model learns the Ambiguous sequences more slowly. This result is qualitatively similar to the empirical result in which a distraction task prevents learning of the Ambiguous sequences. It must be noted, however, that Ambiguous sequence learning does occur in the simulation. For the Unique and Hybrid sequences, the model required over 50 blocks to learn the sequences, while human subjects show learning within 8 blocks of training. It is not clear how the time frames between the model and the human are related, but given that the model required 106 blocks to learn the Ambiguous sequence, it is possible that humans may also show evidence of learning the Ambiguous sequences with more extensive practice.

Simulation 2 -- Hierarchical Organization of Input Sequences

Cohen, et al. proposed that the role of attention in sequence learning is to mediate parsing of sequences containing ambiguous associations. The goal of Simulation 2 was to examine the effects on learning of imposing an hierarchical organization on the input sequence. The associative learning mechanism implemented in Simulation 1 required significantly more training (106 blocks) to learn the Ambiguous sequences than to learn the Unique and the Hybrid sequences (65 blocks). Preorganizing the input patterns into parsed subsequences should allow the associative mechanism to now learn the Ambiguous sequences as quickly as it learns Unique and Hybrid sequences. Because the associative mechanism is already efficient at learning Unique and Hybrid sequences, performance on them should remain virtually unchanged by delineating boundaries in the sequence patterns.

Training Set

The training sets in Simulation 2 were similar to those used in Simulation 1 except that, in Simulation 2, parsing was imposed on the input patterns. Parsing was achieved in two alternative ways. In one condition (Simulation 2.1), context unit activation was set to 0 at the beginning keypress of each cycle of the sequence within a block of 20 cycles. Unique, Hybrid, and Ambiguous sequences were parsed in this way and presented to the model. In this condition, the internal structure of the sequence is left intact, but the beginning of each cycle is delineated by a change in context. Here parsing is achieved by making explicit changes in activation values of the context units, while leaving plan unit values constant. In another condition (Simulation 2.2), which affected only the Ambiguous sequences, parsing was achieved by modifying the value of the plan units at the beginning of each subsequence within the Ambiguous sequence. For example, the sequence 3-1-2-1-3-2 had a plan value of 110 for the 3-1-2 subsequence, and 111 for the 1-3-2 subsequence. During training, therefore, the subplan values alternated as the sequence progressed. In this condition, the Ambiguous sequence is represented as two alternating subsequences. Here parsing is achieved by making explicit changes to the plan units, while leaving the context unit values dependent on model function.

Results of Simulation 2.1

In the first parsing condition, resetting context values to zero, the Ambiguous sequences were learned as quickly as were the Unique and Hybrid sequences (mean blocks 50.8, 60.7 and 57.0, respectively). The three sequence types are not significantly different ($F(2,15) = 1.483$, $p< 0.26$).

An analysis of variance of the number of blocks of training required to learn the sequences in both Simulation 1 (unparsed) and Simulation 2.1 (parsed, zero context) revealed significant effects of sequence type (Unique, Hybrid, or Ambiguous) ($F(2,30) = 7.991$, $p < 0.002$), sequence organization (unparsed vs. parsed) ($F(1,30) = 33.659$, $p < 0.000$), and a significant interaction of sequence type and organization ($F(2,30) = 18.088$, $p< 0.000$). Unique and Hybrid sequences are learned more quickly than Ambiguous sequences (mean blocks = 61.4, 62.5, and 78.4, respectively). Parsed sequences of all three types are learned more quickly than unparsed sequences (mean blocks = 56.2 and 78.7, respectively). The main effects of sequence type and organization, as well as the interaction, are strictly the result of improvement in learning rate for the Ambiguous sequences in the parsed condition over the rate in the unparsed condition. Learning rates for Unique and Hybrid sequences are unimproved by parsing.

Results from post hoc comparisons (Tukey, $p < .05$) of learning rates for the three sequence types in the parsed and unparsed conditions support the hypothesis that unparsed Unique and Hybrid sequences as well as parsed Unique, Hybrid, and Ambiguous sequences are all learned at the same rate (mean blocks = 59.74). Only unparsed Ambiguous sequences are significantly more slowly learned (mean blocks = 106.00). The groupings are as follows:

GROUP A: Unparsed (Unique, Hybrid); Parsed (Unique, Hybrid, Ambiguous)
GROUP B: Unparsed (Ambiguous)

Results of Simulation 2.2

In the second parsing condition, involving alternating subplans in Ambiguous sequences, the parsed sequences were learned in 55.3 blocks. An analysis of variance of blocks to learn the Ambiguous sequences under three types of organization (unparsed, parsed (zero context), and parsed (alternating subplans)) reveals a significant effect for organization ($F(2,15)=24.958$, $p<0.000$). Post hoc comparisons show that the parsed sequences are learned at the same rate, while the unparsed sequences require significantly more exposure to the sequence to learn it. In this case, the particular implementations of parsing are not significantly different. Figure 3 summarizes the results of Simulation 2, displaying the rate of improvement in model prediction as a function of training for parsed and unparsed Ambiguous sequences. The learning rate for both implementations of parsing is similar to that required to learn the parsed Unique and Hybrid sequences, while the learning rate for the unparsed Ambiguous sequence remains at a relatively high level of prediction error.

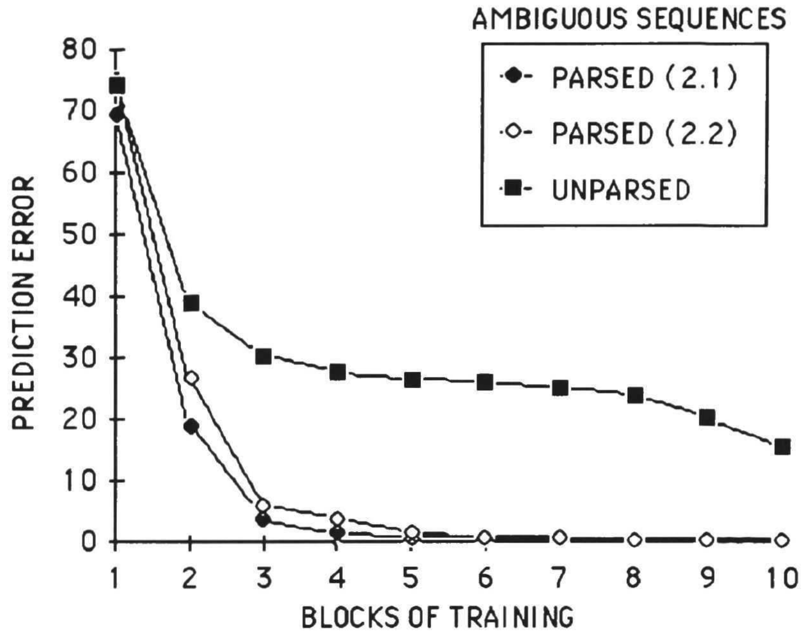

Figure 3. The effect of sequence organization on learning Ambiguous sequences

Discussion

Each particular method of implementing parsing had the same effect on sequence learning. When parsed in either of these two ways, Ambiguous sequences are learned by an associative mechanism at a rate similar to that required to learn Unique and Hybrid sequences. These results agree with the empirical results in which removal of distraction (presumably allowing an attention-dependent parsing process to function) enables, or at least greatly enhances, learning of the Ambiguous sequences.

GENERAL DISCUSSION

The two forms of parsing used in the simulation correspond to two different mechanisms, both of which could be involved in human sequence learning. One form of parsing involves resetting context units to zero at the beginning of each cycle of the sequence while leaving plan values intact. This form of parsing may be seen as mimicking a working memory function. The sequence is represented by activation across the plan units, but learning is facilitated by identifying the boundaries of the sequence. When a cycle of the sequence ends, plan values denote that the same sequence will follow, but the change in context values provides information about where the sequence begins and ends as well as indirect information about the length of the sequence.

The second form of parsing was one in which plan values were altered to represent sequence subparts, while leaving the context units to function uninterrupted. This notion of parsing is somewhat closer to traditional notions of hierarchic representation in which a node stands as a symbol for a more complex and lower order sequence of events. It may be seen as simulating the function of higher level object recognition processes. The alternating subplans represent the final products of a mechanism which has identified the subparts of a sequence whole.

The components of the Jordan model provide further insights into the various computations involved in learning complex motor sequences. The foundation for this learning system is the associative mechanism represented in the model by connections strengths between input patterns and response predictions. A simple two-layer system is fully capable of learning the Unique sequences in these simulations. Hidden units mediate higher level abstractions of input patterns. The hidden units perform an intermediate synthesis of input features. Context units function in a way that is analogous to visuospatial working memory. And finally, plan units reflect the contributions of representational or symbolic processes.

The model used here to simulate sequence learning represents the major components of a sequence learning system. Some of these component processes are dependent on attention, while others are not. It appears that various types of motor sequences require differing levels of interaction with these representation and memory components depending on the structure of the sequence. Unique sequences may require only a strict associative mechanism. Hybrid sequences may require association and synthesis as well as working memory capacity. Ambiguous sequences may need the full system of association, synthesis, working memory, and symbolic representation. These simulations have shown that learning of ambiguous associations are facilitated by higher level processes of organization. Facilitation of learning may occur at the level of working memory by providing cues to pattern boundaries. Facilitation may also occur as the result of an attention-dependent process which parses input sequences into subparts in a hierarchical representation. These effects of different parsing implementations provide insight into the reason that Hybrid sequences are learned by human subjects and by the simulation at a rate similar to that required to learn Unique sequences. The unique elements in the Hybrid sequence may serve to provide cues to boundaries within the sequence. These cues may function at the level of working memory and are independent of attention. This would allow Hybrid sequences to be learned during distraction, as the empirical data show. The Ambiguous sequences contain no clues to pattern boundaries, so they must be analyzed into subparts to be learned. This hierarchical organization process requires attention. The empirical data show that human subjects are able to learn the Ambiguous sequences when attention is not distracted to another task, but do not show evidence of learning the Ambiguous sequence under dual-task conditions. Future simulations will involve examining further the nature of these interactions between sequence structure, attention, and the fundamental operations that underly sequence learning.

ACKNOWLEDGMENTS

We are grateful for the financial and moral support from the Office of Naval Research (Contract N00014-87-K-0279) in the conduct of this research. We have benefitted greatly from discussions with Asher Cohen, Richard Ivry, and Clifford Keele.. We extend our thanks to Jay McClelland for making available a test version of Macintosh software on which these simulations were run. Discussions with Jim Tanaka during the early work on these simulations were helpful and very much appreciated. And thanks to Mike Posner for comments and discussions that are forever thought-provoking and insightful.

REFERENCES

Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 17-30.

Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach* (ICS Report 8604). La Jolla, CA: Institute for Cognitive Science, University of California, San Diego.

Jordan, M. I. (1990). Learning to articulate: Sequential networks and distal constraints. In M. Jeannerod (Ed.), Attention and performance XIII. Hillsdale, NJ: Erlbaum.

Keele, S. W., Cohen, A., & Ivry, R. I. (1990). Motor programs: Concepts and issues. In M. Jeannerod (Ed.), *Attention and performance XIII.* Hillsdale, NJ: Erlbaum.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19,* 1-32.