A Theory of Skilled Memory[1]

James J. Staszewski
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
jjs@cs.cmu.edu

Abstract

A theory of mnemonic expertise is outlined along with findings from initial tests. The expertise belongs to a normal adult (DD) who developed a digit-span of 104 through extended practice. The theory describes how mechanisms consistent with the principles of skilled memory (Chase & Ericsson, 1982; Ericsson & Staszewski, 1989) and identified by analyses of DD's behavior, support his skill. Implemented as a computational model, the theory assumes that distinct knowledge structures mediate both DD's encoding of short segments of trial lists as elaborate, well-structured LTM representations and their retrieval in several recall tasks. Current testing investigates the model's ability to generate *contextual codes*, a class of patterned memory elaborations experimentally shown to improve DD's serial recall (Staszewski, 1990). Given the same lists DD received, it successfully generates over 80% of the contextual codes in his verbal reports. Because successful simulation of contextual codes entails accurate simulation of operations performed by first-order coding mechanisms, results support theoretical assumptions about the knowledge underlying DD's coding operations. The model's overly powerful coding suggests that more stringent architectural constraint must be incorporated to rigorously demonstrate how skilled memory can increase working memory capacity in a normal cognitive architecture and support expertise.

Introduction

How does the human mind adapt to a complex world despite sharp limitations on its computational capacity? Although many architectural constraints on information processing capacity have been identified (Shiffrin, 1976; Simon, 1976), our understanding of their impact upon purposive thought remains uncomfortably sketchy and incomplete. Consider one of the most studied and influential constraints on human cognition, the capacity of short term or working memory (Miller, 1956).

> There is a genuine puzzle here. One the one hand, massive, diverse data demonstrate that STM is limited. On the other hand, no synthetic cognitive systems perform tasks of any scope with such small temporary memory (Newell, 1992).

One solution to this theoretical anomaly asserts that functional working memory capacity is not permanently fixed either to a single memory structure or some small "magical" number of symbols. Rather, as normal adults develop complex cognitive skills, they acquire knowledge structures and memory management strategies that expand working memory capacity. These are the basic tenets of Skilled Memory Theory (SMT), a framework of general principles proposed to explain experts' domain-specific memory advantage and how it contributes to the superior task performance that is the hallmark of expertise.

The empirical foundation for SMT comes from detailed analyses of expert mnemonists (Chase & Ericsson, 1981; 1982; Ericsson & Polson, 1988a, b; Staszewski, 1988b, 1990) and mental calculators (Chase & Ericsson, 1982; Staszewski, 1988a), some of whom developed their exceptional skills under laboratory observation. Results showed that a few common, abstract mechanisms, adaptively tailored to specific contexts, support their capabilities in the same fundamental way: these mechanisms enable extraordinarily efficient storage and retrieval of task-relevant information in long-term memory (LTM) (Ericsson & Staszewski, 1989). Such use of LTM as a cache augments the naturally limited capacity of short-term memory (STM) and increases the amount of information experts can rapidly and reliably access for immediate processing, thereby enhancing information-processing capacity and improving performance.

---

[1] This research was supported by the Personnel and Training Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-85-K-0524.

Although progress in understanding the foundations of skilled memory and its contribution to expert-level cognitive skills is evident, much work remains to establish the validity of SMT[2]. For example, a crucial test of any psychologically-oriented information-processing theory is whether or not it can be expressed as a program that simulates the behavior of human subjects in theoretically significant tasks. Accordingly, the goal of this work is to test the principles of skilled memory by testing a theory of expert performance, whose core mechanisms are implemented according to these principles. Previous studies have determined the function and structure of these mechanisms, but it remains to be seen if they can operate as an ensemble within a system that behaves as an expert mnemonist does on tasks testing his memory.

DD was an undergraduate who practiced the digit-span task 4-5 days per week under laboratory observation, adapting the mnemonic strategy given him by Chase & Ericsson's (1981) laboratory-made mnemonist, SF. Over 4.5 years, DD's span grew from 7 to 104 digits[3]. Explaining his unprecedented digit-span is a major goal of this theory, but studies indicate that DD's skill is far broader and very complex. Although it is restricted to remembering number sequences, DD demonstrates exceptional memory in a variety of tasks including backward and forward serial recall, incidental and delayed serial recall, free recall, probed recall, and Luria's Matrix recall (Staszewski, 1988, 1990). This theory is intended to account for many theoretically interesting phenomena related to his performance over this range of tasks. Initial tests of the model instantiating this theory examine its ability to simulate a particular phenomenon related to DD's list encoding. This is his generation of *contextual codes (CCs)*, a class of patterned memory elaborations inferred from verbal protocols taken after digit-span trials and experimentally shown to speed his serial recall (Staszewski, 1990).

Overview of the Theory

This theory's fundamental assumption is that DD has a normal memory system with two principle structures, STM and LTM, whose functional parameters fall within normal, well-documented

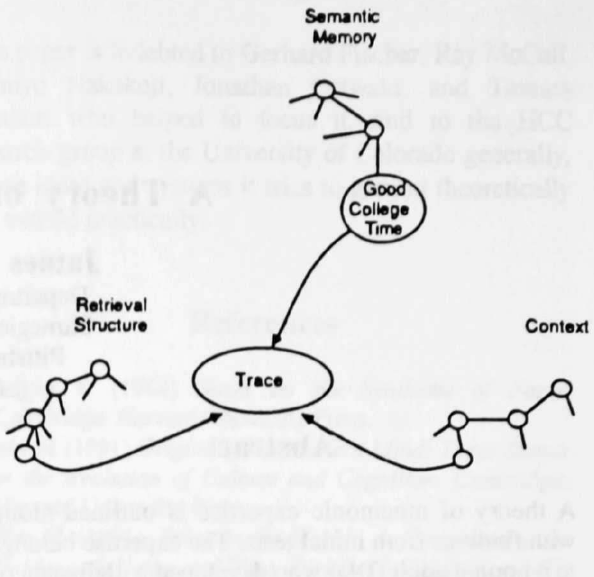

Figure 1. Schematic structure postulated for DD's memory representations for digit groups.

ranges. His memory differs from that of unskilled people in that a specific body acquired knowledge enables him to quickly recognize and encode short digit sequences as well-organized, multi-component LTM representations, whose content and organization support multiple systematic retrieval strategies. These strategies are mediated by same knowledge components that govern his list encoding processes. The structure and general content of the LTM codes is illustrated schematically in Figure 1, along with the mechanisms that contribute information both to their formation and retrieval.

Mechanisms Supporting Skilled Memory

A cornerstone of DD's skill is a hierarchically-organized taxonomy of 12 semantic categories. He uses this knowledge to recognize and encode any sequence of 3 or 4 digits as an organized, meaningful chunk. Nine categories encode digit groups as times for running events ranging from 1/4 mile to 10 mile distances. For example, DD codes sequences like 359 and 8201, respectively, as "a one-mile time, three minutes, fifty-nine seconds, just under the big [four-minute mile] barrier" and "eight minutes, twenty point one seconds, a fast college two-mile." The other categories code digit groups as either ages (391, "just under forty years old) dates (1776, "Revolutionary War started," or numerical patterns (987, "sequence descending from nine"). DD's categorization of digit-groups is very reliable and has been simulated with 98% accuracy (Staszewski, 1990).

---

[2] Gobet's study (in this volume) suggests that SMT generalizes to chess experts' memory skill.

[3] DD has recalled longer random digit lists perfectly, however Woodworth's (1938) metric is applied here: digit-spanis defined as the longest list that can be perfectly recalled on 50% of test trials.

The theory also assumes that systematic activation of this taxonomic structure generates cues for retrieving stored digit groups. This accounts for DD's ability to accurately recall digit groups from either a single list or from all the lists of a practice by categories (Staszewski, 1990).

DD's *retrieval structure* is another essential component of his skill that is used to store and retrieve the chunks he encodes. As list presentation proceeds on a typical digit-span trial, processes "traversing" this abstract structure govern parsing of the list into a predetermined sequence of digit groups. These processes generate an index or "address" for each semantic code, analogous to a call number on a library book; each address constitutes an access route to each chunk for ordered retrieval because the addresses also code ordinal relations among DD's chunks. For serial recall, another traversal or (re)activation of the retrieval structure then regenerates the same addresses, which now function as retrieval cues that access each chunk in the right order. Chronometric analyses of DD list encoding and serial recall operations, his pattern of latencies in a probed recall task, and his verbal reports provide converging evidence from which the form and function of DD's retrieval structures have been inferred (Chase & Ericsson, 1982; Staszewski, 1988b).

Still another mechanism is inferred from another coding regularity found in DD's verbal reports: he consistently notices a variety of abstract symbolic patterns and explicitly codes them with generic labels. For example, he consistently notices when a subset of digits in a group sum to another (i.e., 752, 3012 - both coded "add 'em ups") or that pairs of digits in a group represent quantities that differ by a multiple of 10 or 1 (e.g., 8939, 7675 - coded as "10-apart" and "1-apart" respectively). Alternatively, if two consecutive groups receive the same semantic code, DD typically elaborates his representations for these groups with the phrase "back-to-back [semantic code]s." The "back-to-back" label also applies to consecutive groups previously elaborated as "add 'em ups" or "X-apart". Frequently, if two remote digit-groups[4] receive a very similar semantic interpretation (e.g., 462 and 460, "two fast 1/4 mile times"), whether they are from the same list or from different lists within a test session, the elaboration "X-tenths apart" often explicitly links them. The label for such elaborations as these is contextual code (CC). Examples of more CC's found in DD's verbal reports and a more complete description of the variety of CCs observed can be found in Staszewski (1990).

Why the label contextual? Whereas generation of a semantic code and a retrieval structure address are obligatory operations for each digit group in a list, a CC may or may not be created for a given digit group, depending upon contextual variables. These include the contents of a particular list, the contents of lists presented on earlier trials in a test/practice session, and the amount of time DD has to detect a CC and integrate it with the semantic code and address created for a particular digit group. This time factor is a non-trivial constraint on CC formation considering the one-digit/sec list presentation rate used in practice and most experiments. It follows that specific CCs don't appear with the regularity that marks DD's semantic coding and list parsing. However, examination of verbal reports from dozens of practice sessions each with multiple trials reveals regularities that beg description and explanation. They are the basis for postulating that DD's knowledge base contains knowledge structures with distinct category labels that are used to detect and encode a variety of contextual elaborations.

Regarding the psychological reality of CC's, an experiment investigating both their predictability and effect upon DD's memory has shown that (a) his generation specific CCs can be predicted from list contents, and (b) that his encoding of CCs increases the speed of his serialrecall (Staszewski, 1990). Moreover, verbal protocols taken during DD's serial recall show that contextual codes are sought for intentionally when his initial attempt to retrieve a digit group via retrieval structure cues and a semantic code fails. CCs then retrieved in such circumstances are used in a backup retrieval strategy that has a clearly constructive character.

Implementation

The mechanisms and processing assumptions outlined above are instantiated as a production-system model written in OPS5. Space constraints limit description to the features of the model directly relevant to the evaluation described shortly.

Input for all test trials consists of a parameter indicating list length and then a digit list of that length. Output for the current test consists of the digit groups into which the list is parsed, the semantic code assigned each, and any label of any CC the model detects related to a particular group.

DD's semantic coding structure and retrieval structure are represented as hierarchical list structures, whose nodes contain information that is activated by a simple node-traversal operator and consequently loaded into working memory. The information consists of semantic features related to coding categories in the former case, and abstract attribute-values representing retrieval structure addresses for the latter. Data structures (i.e., schemas) for addresses and semantic codes are used to integrate and represent these feature sets as chunks in working memory for

---

[4] Groups separated by one or more intervening digit groups.

each digit group within an input list. These chunks are the elements that get attached to the generic memory structure shown in Figure 1.

Sixty-eight productions find CCs and output the appropriate generic labels. The condition elements of each always includes information about either group addresses, or semantic features, or sometimes both (depending on the type of CC). This makes accurate detection of CCs dependent upon accurate simulation of first-order coding operations (retrieval structure addressing and semantic coding).

Coding operations are ordered in the following way. First, the location for each digit-group is encoded, because the same processes that govern list parsing generate group addresses. After the registration of the second digit of each group, a sorting procedure operating like a discrimination net uses the two digit values to select the appropriate semantic category label, which is then output. In a few cases, an additional test based on the value of the third digit is necessary for assignment of a semantic code. Only after these obligatory coding operations are completed can productions for detecting contextual codes fire.

Testing Procedures

The model was given the same lists presented to DD in Staszewski (1990) as input for each trial. Trials were presented in blocks of 6, corresponding to the number of trials and ordering used in each of the eight separate sessions in which DD was tested. The model's list parsing, assignment of semantic codes to digit groups, and generation of CC's were all compared to the same information found in DD's verbal protocols for each test list.

Scoring involved matching the model's coding against information in DD's verbal report for each list. Relatively stringent criteria were used. A CC match was scored if the model met all three of the following conditions: (a) the model grouped digits in the same way as DD where he reported a CC, (b) relevant digit group(s) were labeled with the same semantic coding category that DD used, and (3) the model generated the same CC label as DD reported.

Results

The data tabled below show the fidelity with which the model's coding simulated DD's. Data aggregated across all sessions and code types (Row 1) show that the model generated 82% of the 345 CCs in the verbal protocols. The model's list parsing and assignment of semantic codes coincided with DD's nearly perfectly (over 99% of the time).

The model also generated many CC's where DD reported none, as the *false positives* column shows. Closer analysis of model's performance revealed that many of these errors occurred where patterns of information were present that generated *multiple*, redundant CCs for the same digit group or set of groups. For example, CCs for triplets of semantic codes (e.g., "three consecutive one-mile times") entail minimally the presence of two lower-order CC's (2 consecutive "back-to-back" pairs of miles, and one "back-to-back" set of "back-to-back" CCs). Intra-group and between-group CC's based on digits sometimes further increase the number of CCs in such a situation.

Assuming that (a) one explicitly coded CC is sufficient to enhance memory for a digit group and (b) that time constraints of list presentation(to which the model was not subject) make exhaustive coding of elaborations implausible, a conditional analysis of false-positive was carried out. This analysis simply discounted redundant CCs in local list regions where one CC match had been found. As the conditional data show, the number of false positives decreases by a factor of 3.6 under these assumptions.

Finer-grained analyses based on subcategories of contextual codes yield the data shown in rows 2-5. Analyzed by types of CCs, results show that the model is most accurate (96%) at detecting CCs that associate adjacent digit-groups on the basis of pairs of repeated digits or symbolic codes. The hit rate for the remaining three types hovers around a respectable 75%. Analysis of false-positives by CC type localizes and explains the model's excessive contextual coding power. Matches well exceed false positives for the first two CC types. However, the number of false positives for CCs linking remote digit-groups found within the same list nearly equals

Table 1. Model's identification of DD's contextual codes. Table entries represent frequencies.

Analysis	Matches	Misses	False Positives	Conditional False Positives
CCs, all types, all sessions	282	63	202	56
CCs linking digits, within groups	84	31	49	11
CCs linking adjacent groups	118	5	38	9
CCs linking remote groups within lists	48	17	41	10
CCs linking remote groups across lists	32	10	74	26

the number of matches. False-positives more than double the number of matches for CCs linking digit-groups from different lists. Conditional analyses reduce their number from 65% to 81% for the different CC types, but the frequency of false positives, particularly for CC's that span lists, pinpoint significant limitations of this model of DD's coding processes.

Discussion

The model's successful generation a substantial proportion of DD's CCs give it credibility as a first approximation of the knowledge and processes that support his mnemonic encoding skill. This conclusion takes into account the fact that accurate simulation of DD's list parsing, retrieval structure addressing, and semantic coding processes are prerequisites for simulating his generation of CC's.

At the same time, problems with the model indicate that further development will be necessary to produce a theoretically sound account of DD's skill. The excessive power demonstrated by CC mechanisms in this study, a phenomenon readily interpretable in terms of model limitations, both motivates and guides future modification.

First, with no time constraints on encoding, the model can exhaustively search for, find, and code all the CC patterns available within a session. On the other hand, temporal constraints related to conventional list presentation rate used in DD's practice (one digit/sec.) should limit the amount of contextual coding he can do, if his STM is normal, that is, time-based. Results of the conditional analyses, which eliminate generation of redundant CC's and limit the number of CCs per digit-group to one, suggest that the current model may be able to incorporate this constraint without substantial reconfiguration.

Second, regarding false-positives that remain even after redundant CCs are ignored, the type of CC which shows the largest number is that which relates groups across lists. Even though some digit groups that the model links with CCs are separated by as many as four intervening lists, their detection poses no problem for a system that doesn't forget and is not hampered by time constraints on search and encoding. DD's memory is fallible, however, as his digit-span of *only* 100 or so digits shows. These observations indicate that some provision for forgetting must be made in future versions.

In conclusion, results of this work support key theoretical assumptions about the knowledge and processes that support DD's skilled memory. They also show that reasonable architectural constraints must be incorporated in future versions to test the proposed theory with appropriate rigor.

Acknowledgments

Thanks to Fernand Gobet, Bob Siegler, and Herbert Simon for their helpful comments on this paper.

References

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Erlbaum.

Chase, W. G., & Ericsson, K. A (1982). Skill and working memory. In G. H. Bower, *The psychology of learning and motivation,Volume 16*. New York: Academic Press.

Ericsson, K. A., & Polson, P. G. (1988a). An experimental analysis of the mechanisms of a memory skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*,305-316.

Ericsson, K. A., & Polson, P. G. (1988b). Memory for restaurant orders. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Gobet. F. (in this volume). A computer model of chess memory.

Miller, G. A. (1956). The magical number seven, plus or minus two. *Psychological Review, 63*, 81-97.

Newell, A (1992). SOAR as a unified theory of cognition: Issues and explanations. *Behavioral and Brain Sciences, 15*, 464-492.

Shiffrin, R. M. (1976). Capacity limitations in information processing, attention, and memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes, Volume 4*, Hillsdale, NJ: Erlbaum.

Simon, H. A.(1976). Neural mechanisms of learning and memory. In M. R. Rosenzweig and E. L. Bennett (Eds.), *The information-storage system called 'human memory'*. Cambridge, MA: M.I.T. Press.

Staszewski, J. J.(1988a). Skilled memory in expert mental calculation. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Staszewski, J. J. (1988b). The psychological reality of retrieval structures: An investigation of expert knowledge. *Dissertation Abstracts International, Volume 48, 2126B*. Doctoral dissertation, Cornell University, 1987.

Staszewski, J. J. (1990). Exceptional memory: The influence of practice and knowledge on the development of elaborative encoding strategies. In W. Schneider & F. E. Weinert, *Interactions among aptitudes, strategies, and knowledge in cognitive performance*. New York: Springer-Verlag.

Woodworth, R. S.(1938). *Experimental psychology*, New York: Holt, Rinehart, & Winston.