MACHINERY OF HUMAN MEMORY

Richard C. Atkinson

My goal today is to describe what is called, in the literature of the psychological and cognitive sciences, the Atkinson-Shiffrin theory of memory. The theory was set forth in a series of papers in the 1960s **[Figure 1]**. The principal paper was published in 1968 entitled "Human Memory: A Proposed System and its Control Processes." My co-author was a young graduate student at Stanford University named Richard Shiffrin. I'll have more to say about Rich later. In 2019, the journal *Memory and Cognition* devoted a special issue to the theory entitled "Recognizing Five Decades of Cumulative Progress in Understanding Human Memory and its Control Processes Inspired by Atkinson and Shiffrin (1968)."

The theory is in one sense the culmination of work dating back to the earliest days of psychology as a field of study. William James' writings on memory are a leading example, but there is a long history of speculation about the mysteries presented by memory and ideas to explain them. We used many of these ideas as a starting point. The goal was to organize them as a *general theory* from which one could derive *formal models* to predict an individual's performance on various memory tasks. I use the term *formal model* to describe both mathematical and computational models. Simply put, we wanted to place the theory on a quantitative basis.

It was an auspicious time. By the late 50s and early 60s experimental psychology was moving in new directions. What is usually called the cognitive revolution had (among other things) stimulated a strong interest in the machinery that drives the many ways memory operates. Modeling was making it possible to explain findings on memory more precisely, using quite simple assumptions.

The experimental data and the predictions our models generated turned out to be consistently accurate and technically rigorous. This gave the theory considerable scientific validation. I also suspect it has been successful because we were able to organize the theory's basic concepts in such a way that they could be employed by other scientists to generate alternative models. The Atkinson-Shiffrin theory has evolved in light of new research and survived a number of competing theories and critical challenges. More than half a century later, it is still going strong.

HUMAN MEMORY: A PROPOSED SYSTEM AND ITS CONTROL PROCESSES[1]

R.C. Atkinson and R. M. Shiffrin

STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Figure 1

Before launching into the details, I need to provide some background regarding the types of experiments that are the focus of the theory. Since the middle of the 19th century scientists in Germany and elsewhere in Europe, and later in the United States, have been conducting well-controlled laboratory experiments on the psychology of memory: how individuals memorize information and then retrieve that information at a later time--seconds, minutes, or days later and sometimes years. Many of you have probably been subjects in one or another of these experiments. Anyone who took introductory psychology at a major university usually had an opportunity to participate in several such experiments.

An example of a very simple experiment is what's called *free recall.* The subject first sees a long list of unrelated words presented one at a time. After all the words have been presented, the subject must then recall them in any order (hence the designation *free recall*). The result of principal interest is the probability of recalling each word as a function of its place in the list **[Figure 2]**. Plotting this function yields a U-shaped curve as illustrated in the top curve of Figure 2 (for a list of 15 words presented at a rate of one per second). The increased level of recall for the first few words presented (on the left of the graph) is called the *primacy effect*; the large increase for the last eight to twelve words (on the right) is called the *recency effect*. I'll return to this example later.

Over the years, a variety of experiments have been devised probing every aspect of memory. As the data accumulated for a specific memory task, investigators proposed explanations for what was taking place. These explanations were usually stated in qualitative terms; they provided some insights into the memory tasks being investigated, but there was little generality from one to the next.

The general theory, the term used in our 1968 paper, is formulated as an *information processing system.* It postulates a distinction between the *memory system* itself and the various *control processes* that manage the flow of information within the system **[Figure 3]**. The general theory is specified in sufficient detail that stochastic models can be built to predict a subject's performance in a wide range of experiments.

The system has three separate memories: the Sensory Register (SR), the Short-term Store (STS), and the Long-term Store (LTS). The SR is a short-lived store for sensory inputs, measured in milliseconds. I'll say more about it later. The STS is of limited capacity. Its content is continually changing, nothing is stored

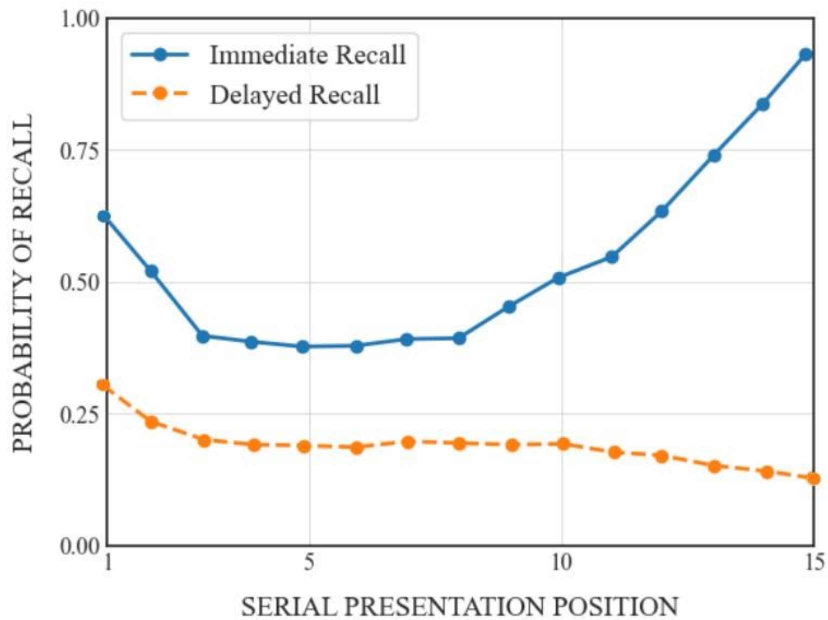


Figure 2

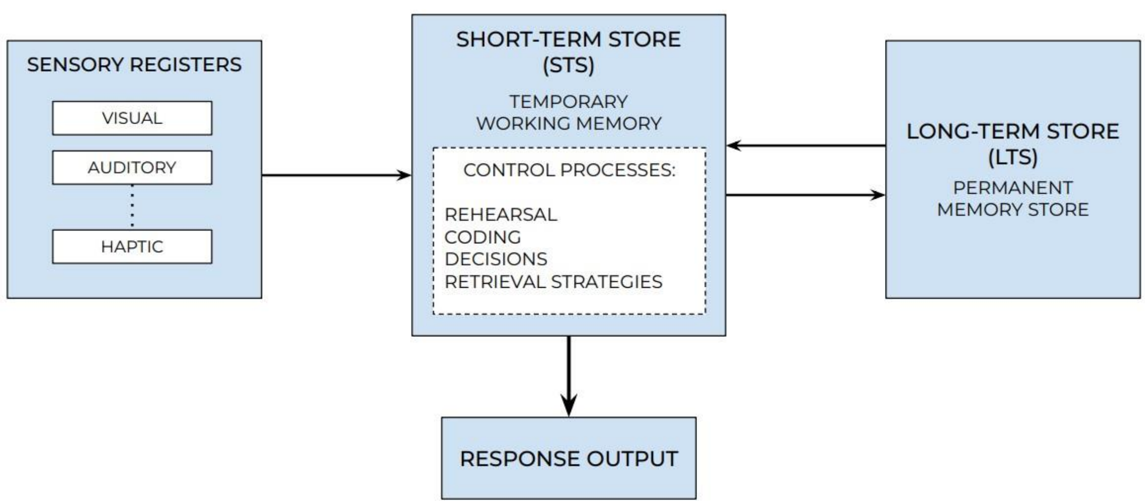


Figure 3

there on a permanent basis. If you like, think of your conscious experience as part of the contents of STS. In contrast to STS, LTS is virtually limitless and provides a relatively permanent repository of information. Stored in LTS is information about episodes that occurred over a lifetime, the knowledge needed to understand and speak a language, and all other information available to us from memory.

Information from the environment is processed by the various SRs (vision, hearing, etc.) and selectively entered into STS. Information once entered into STS can lead to the retrieval of related information in LTS. The retrieved information helps interpret what is currently in STS and determines what information should be transferred to LTS. A strong assumption of the theory is that the storage of information in LTS must occur via STS. STS is the gatekeeper for what is stored in LTS. What drives the whole system is a set of control processes that determine what stimuli are attended to, whether or not they are encoded in STS, and if encoded will be transferred to LTS.

To give you some sense of the system in operation, let me describe a case study of an individual with amnesia of a unique kind. In 1953, a young man known by his initials HM (to protect his identity) underwent an experimental operation involving the removal of a brain structure called the *hippocampus.* The surgeon hoped that the operation would relieve HM's chronic seizures; it proved less than successful. However, the operation left him with profound amnesia of a very special kind. When he was introduced to a new person, he could carry on what appeared to be a normal conversation. However, if he saw the same person again (an hour or a day later) he would have absolutely no recall of having met and talked with the person before. In conversation, he had no trouble recalling events that occurred prior to the operation such as World War II, the Great Depression, or the presidency of Franklin Roosevelt. But events that occurred after the operation (the Vietnam War, Nixon's presidency, or the demise of the Soviet Union) all came and went without leaving a trace in LTS. For HM, short-term memory continued to work reasonably well, as did the retrieval of information from LTS placed there prior to the operation. However, after the operation HM lost all ability to transfer new information to LTS. This, and other neurological evidence, indicates that the hippocampus is critical in the transfer of information from STS to LTS.

One of the first applications of the general theory, called "the buffer model," was to the free recall task described earlier. The model is simpleminded but powerful. To this day, the buffer or some variant is a component of virtually every model of the general theory. For the free recall task, the basic idea is that a subject will set up a buffer in STS that can hold and maintain a fixed number of items. At the start of the presentation of a list, the buffer is empty; successive items are entered until the buffer is filled. Thereafter, as each new item enters the buffer, it replaces one of the items already there. The items that are still in the buffer when the last item is presented are the ones that are immediately recalled, giving rise to the recency effect. The transfer of information from the buffer to LTS is postulated to be an increasing function of the length of time an item resides in the buffer. Since items presented at the beginning of the list enter an empty or partially empty buffer, they remain longer than later items. This extra time in the buffer for the first few items on the list gives rise to the primacy effect.

Simply stated, (1) an item still in the buffer at the time of the test will be recalled immediately, and then (2) an item no longer in the buffer will be recalled as an exponential function of the time, t, it resided in the buffer. Let r denote the size of the buffer. In experiments of this sort, r is typically 7 plus or minus 2. The number 7 is sometimes called the "magical number" because of its frequent occurrence in memory research. Let Θ be the transfer rate of information from STS to LTS. Then the probability of recalling an item in LTS that resided in the buffer for time t is $1 - \exp(-\Theta t)$.

Let me now return to Figure 2. In a typical free-recall experiment several lists are presented to a subject for immediate recall, all lists involving different words. The U-shaped curve is for *immediate recall* as described earlier. In some experiments, at the end of a session, the subject is asked (without prior warning) to recall all words from all lists in any order. The bottom curve is for *delayed recall* and is the probability of recalling a word as a function of its serial position within a list. The immediate recall curve represents retrieval from both STS and LTS. The delayed recall curve is retrieval from LTS; it can be regarded as a direct measure of the time an item resides in the buffer. The buffer model provides an excellent fit to these data, once Θ and r have been estimated.

The parameters Θ and r can be manipulated experimentally by varying different aspects of the free-recall task. For example, using auditory versus visual stimuli, the rate at which items are presented, list length, age of the subject, and many other variables. With remarkable accuracy, the buffer model provides an excellent fit for a wide range of experiments.

Another example of a free recall experiment involves what are called paired associates, a stimulus paired with a response. For example, pairing a nation's flag as a stimulus with the nation's name as a response. **Figure 4** presents data from such an experiment, with a "number" as the stimulus and a "color" as the response. As in the free recall task, a list of paired associates is presented one at a time and then the subject is tested by presenting one of the stimuli. The correctness of the

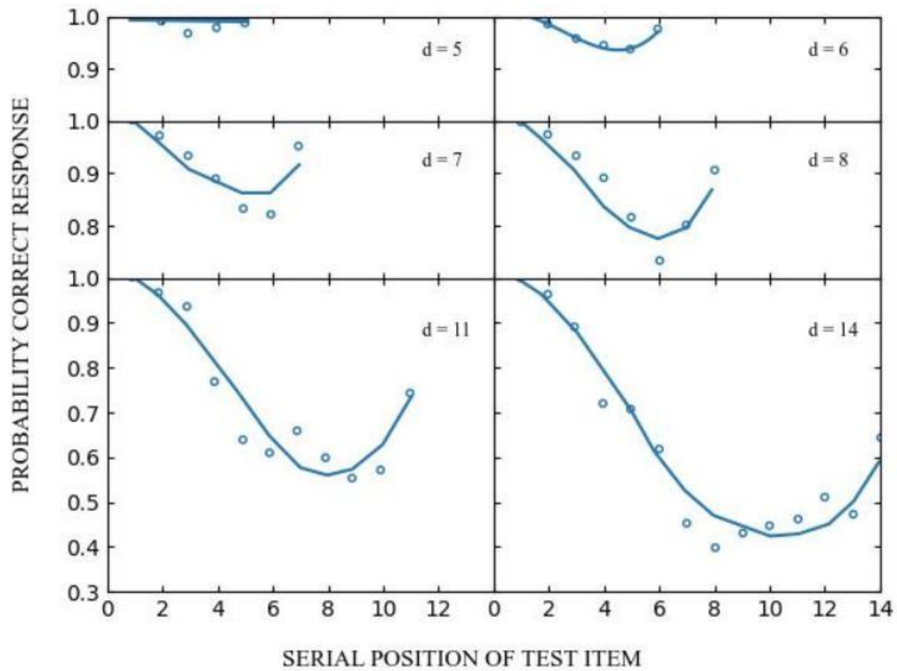


Figure 4a Probability of a correct response as a function of list position and display size. Observed values are denoted by circles, and theoretical values by the solid lines. (x^2 = 44.3 based on 42 *df*.)

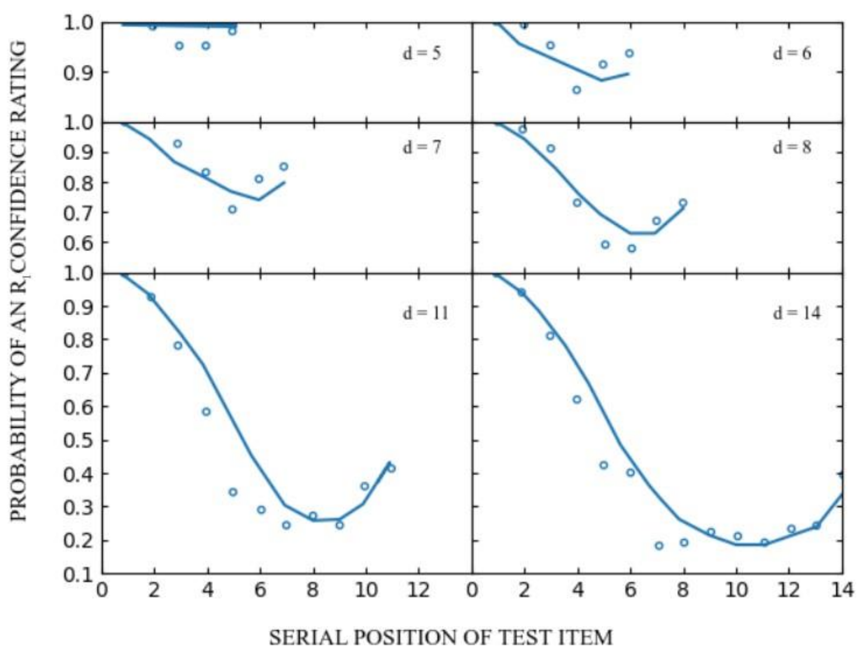


Figure 4b Probability of a confidence rating R_1 as a function of list position and display size. Observed values are denoted by circles, and theoretical values by the solid lines. (x^2 = 111.9.)

response and the subject's confidence in the response are recorded. The predicted and observed results are presented in Figures 4a and 4b.

I am now going to make a rather sweeping claim. The general theory, with the buffer model in various configurations, could explain most of the extant data on memory collected prior to our 1968 paper. To make this claim more plausible, note that we regard an earlier theoretical development, Stimulus Sampling Theory due to William Estes, as a special case of the general theory and a necessary prerequisite for our work. I was particularly pleased that the theory provides a beautiful account of Ebbinghaus' data first reported in 1885. His work on serial learning is generally credited with establishing the field of experimental research on memory.

After publication of the 1968 paper, a number of studies were reported that claimed to disprove the general theory. The problem was that some critics viewed the buffer model as a statement of the general theory, rather than a special case. Responding to the claims usually required a more sophisticated control process. For example, a control process that employed a more complex encoding scheme. Or a control process that allowed the subject's attention to selectively focus on the more important sensory information. The subject can always modify the control process and may do so even given minor changes in the experimental procedure. And some control processes that initially required a subject's attention can become automated, if frequently used.

However, the most significant finding from our earlier research was the variety of ways STS can be deployed. It is a memory of limited capacity, but it is key to how the system operates. It determines what information is transferred to LTS. At any time, it can decouple the memory system from the distractions of the outside world. Moreover, because information can be maintained in STS on a temporary basis, it often serves as a primary memory in situations where new information once used is no longer needed and allowed to decay.

As another example, consider the following not uncommon experience. You are driving to your office for an important meeting. During the drive your mind is focused on the upcoming meeting. As you arrive at the office, you suddenly realize you have no memory of the drive itself. Your STS was simultaneously engaged in two tasks. Subconsciously monitoring your driving to alert you to any untoward events, and at a conscious level planning for the meeting.

In the 1968 paper, we introduced the term "working memory" because STS is where the memory system works its magic. It can create new memory traces, but it can also combine these traces with others already in LTS. For example, the arrival of a close friend will immediately retrieve a strong memory trace with information about her name, age, and family history that has been added to the principal trace over repeated exposures. In addition, there may be other traces that have not been linked to the principal trace. Only with an extended search and a better probe can they be retrieved. Contrast this example with an individual you met once years ago. You may have several memory traces of her that have never been linked together. When meeting her again, you may retrieve a trace that causes you to immediately recognize her, but only later (possibly in the middle of the night) retrieve another trace that lets you recall her name and where you met her.

The last 50 years have witnessed an extraordinary period of experimental research on memory and cognition. A wide array of phenomena have been discovered using clever new types of experiments. In order to explain these developments in terms of the general theory, it has been necessary to elaborate components on the memory system. Research findings on perception and visual memory required a more complex SR. The SR played a minor role in the beginning; indeed some of the early models did not even include it. Now the visual SR, in particular, plays a key role in many applications.

The encoding process that produces a memory trace also had to be elaborated. The immediate sensory input of an event had to be supplemented by related information brought to mind by the event. The way an event is encoded, beyond infancy, depends on relevant information accumulated over a lifetime. Further, information about the "context" in which an event occurred proved to be more important than initially expected. Not just context as defined by the physical surround where the event occurred, but also factors like, for example, mood and motivation. How an event is encoded can be a very complicated process influenced by many variables.

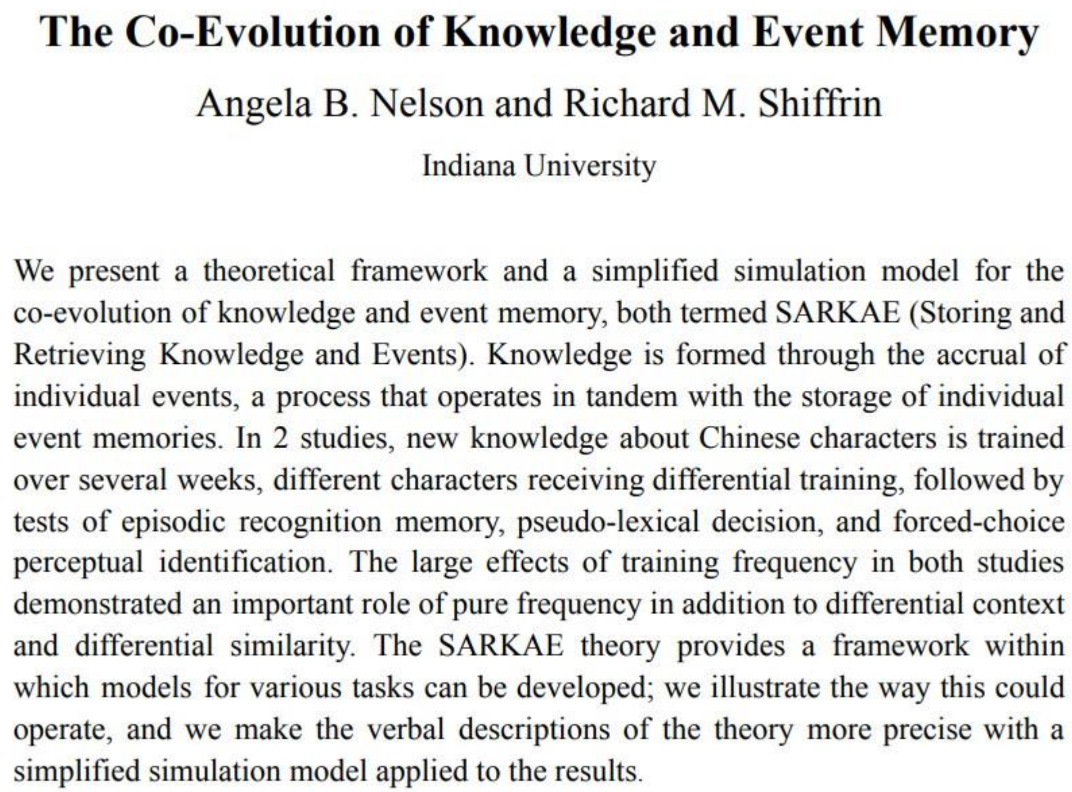
Critical to all of these developments has been the search and retrieval process for LTS. In the early models, we used a simple exponential learning function to approximate the search and retrieval process. As the experiments increased in complexity, it soon was evident that a more competent set-up was needed. For example, research demonstrated that the sheer act of retrieving an event can change our memory of that event. So began an odyssey that continues to this day. A series of models have been developed, each more robust than the last. The basic idea dates back to the 1968 paper. The subject assembles a *retrieval probe* in STS and via search attempts to match the probe to memory images in LTS. Traces are retrieved on the basis of goodness-of-fit between the trace and the probe, with the fit determined by a Bayesian decision. The discovery of several phenomena would not have occurred, except that they were predicted a priori by the model. As an aside remark, the Freudian concept of *repressed memories* can be explained by the general theory as an inability of the subject to generate an appropriate retrieval probe.

Recent developments of the general theory are illustrated by a model called SARKAE (Storing and Retrieving Knowledge and Events). The model is designed to explain the co-evolution of knowledge and event memories. The essential idea is that knowledge in LTS is created by the gradual accrual of individual event memories, whereas the encoding of an event is affected by knowledge already in LTS. These interdependent processes create a feedback loop in which knowledge and event memories grow in concert over a lifetime. The model is a brilliant application of theory and a major scientific achievement **[Figure 5]**.

The related experiment involved learning a collection of Chinese characters over a three-week period. Each subject was trained on five different tasks and given over 12,000 study-test trials to track the accuracy and latency of responses over the course of the experiment. The model for the experiment had a memory trace specified as a vector of 432 features **[Figure 6]**. I mention these details to give you some sense of the size and complexity of the undertaking.

SARKAE is a preview of what research on human memory and learning will look like in the future: large-scale experiments running subjects over an extended period of time. The data from these experiments will be best understood via mega computer simulations of models. As yet, we are not prepared to formulate models for a substantial body of knowledge such as, for example, Euclidean geometry, even though it is a beautiful and well-ordered subject matter. One realistic possibility at this time is the acquisition of initial reading skills by young children. I had considerable experience at Stanford University with a computer-based reading curriculum for grades K to 3 **[Figure 7]**. The types of instructional tasks we used were similar to the tasks used in the SARKAE study. Further, from a linguistic perspective, building a suitable memory trace for a reading program would not be much more difficult than for SARKAE. Modeling the acquisition of reading skills by young children would be a wonderful accomplishment. Moreover, such a model could be used to design a curriculum that optimized the acquisition of reading skills.

Future developments of the general theory could be greatly facilitated by advances in the neurosciences. If we knew more about the neural circuitry of a memory trace, it would provide constraints and guiderails for future models. I also


The Co-Evolution of Knowledge and Event Memory

Angela B. Nelson and Richard M. Shiffrin

Indiana University

We present a theoretical framework and a simplified simulation model for the co-evolution of knowledge and event memory, both termed SARKAE (Storing and Retrieving Knowledge and Events). Knowledge is formed through the accrual of individual events, a process that operates in tandem with the storage of individual event memories. In 2 studies, new knowledge about Chinese characters is trained over several weeks, different characters receiving differential training, followed by tests of episodic recognition memory, pseudo-lexical decision, and forced-choice perceptual identification. The large effects of training frequency in both studies demonstrated an important role of pure frequency in addition to differential context and differential similarity. The SARKAE theory provides a framework within which models for various tasks can be developed; we illustrate the way this could operate, and we make the verbal descriptions of the theory more precise with a simplified simulation model applied to the results.

Figure 5

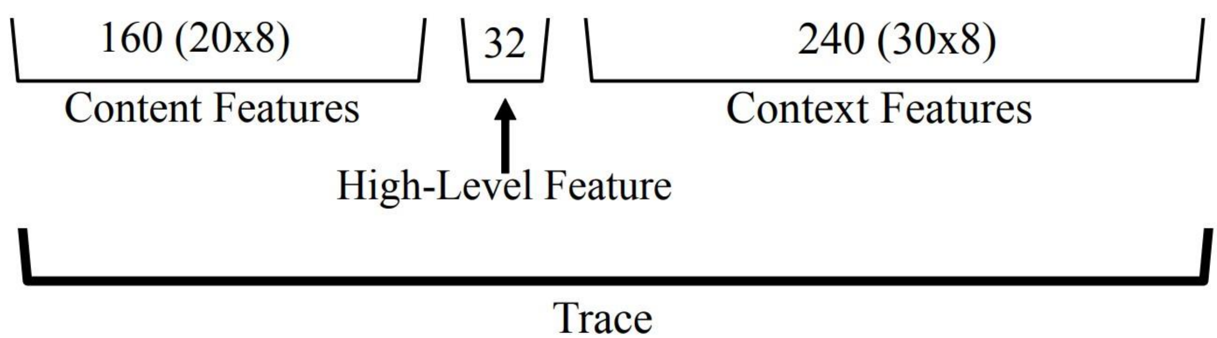
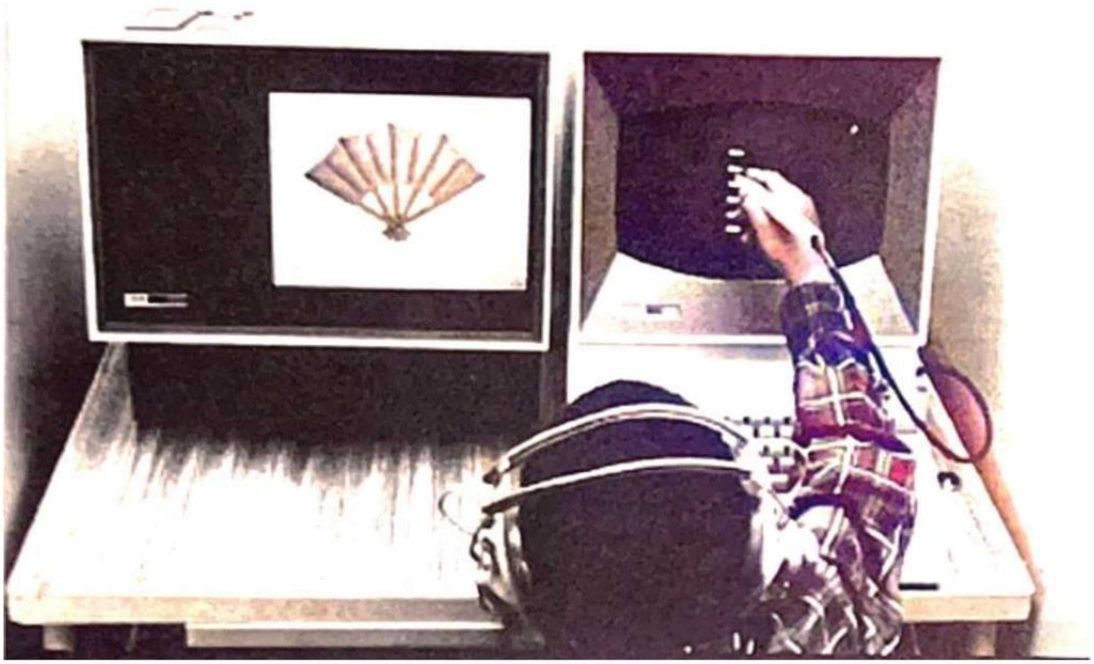


Figure 6



I
Letter
Identification

II
Sight - Word
Recognition

III
Spelling
Patterns

IV
Phonics

V
Spelling

VI
Word
Comprehension

VII
Sentence
Comprehension

Figure 7

am hopeful that developments in AI will identify some heuristics or algorithms humans use in problem solving that have escaped our attention. I should note that John McCarthy, the founder of AI, and my research group both shared a PDP-1 computer at Stanford. A fantastic computer for its time. Most of the early research on the general theory was carried out on the PDP-1, as was John's early work on AI.

By now, you must realize that I am very much smitten with STS. Everything that happens revolves around her. If you want to understand memory, you must get to know STS. Another observation about STS is based on my own metacognition. During certain stages of sleep, STS will turn to a problem that has arisen during the day. She will search deep into LTS for memory traces that contain information possibly relevant to the problem. If several memory traces are identified, they can be mixed together and analyzed. Analysis of this mix may reveal information that was not evident when analyzed individually. The "eureka" experience. An analogy: You have a blue towel and a gold towel. Place them in a washing machine and turn on the switch. Out come two green towels. If reification occurs, STS can create a new memory trace, a trace formed in the absence of external stimulation. The difference between a created memory trace versus an acquired trace reflects the absence or presence of external stimulation. This is an example of STS's creativity that enables her to formulate a theory about the outside world.

In my remarks, it has not been possible to recognize the many people who have contributed to the development of the theory. As you may know, my research career ended when I became director of the National Science Foundation in 1977. Since then, Rich Shiffrin and his many students and postdoctoral fellows have been key to everything that has happened. He has had a remarkable scientific career and received every honor the field has to offer. I am pleased to say that in addition to being a collaborator, he has been a good friend. Finally, let me note that I recently reviewed the manuscript of a book entitled "Human Memory: The General Theory and Its Various Models," to be published by Cambridge University Press. The author is Kenneth Malmberg, a professor at the University of South Florida and a former postdoctoral fellow of Rich Shiffrin's.