Applications of the spacing effect to human learning and memory

James Roller Michaelis

University of Colorado Boulder

Follow this and additional works at: http://scholar.colorado.edu/csci_ugrad

Recommended Citation
Applications of the spacing effect to human learning and memory

Submitted to the University of Colorado at Boulder as part of the requirements for the degree

Bachelor of Science in Computer Science

By: James Michaelis

Supervised by:
Professor Michael C. Mozer
Dept. of Computer Science, University of Colorado at Boulder

Defended on:

May 14th, 2007

Committee Members:

Professor Michael C. Mozer
Dept. of Computer Science, University of Colorado at Boulder

Professor Alice Healy
Dept. of Psychology, University of Colorado at Boulder

Professor James Martin
Dept. of Computer Science, University of Colorado at Boulder
CONTENTS

1 Introduction
1.1 The Spacing Effect – Overview .................................. 3
1.2 The Spacing Effect in Experimental Literature ........ 3
1.2.1 Glenberg (1976) ............................................. 3
1.2.2 Young (1971) .............................................. 5
1.2.3 Pavlik & Anderson (2005) ............................... 6
1.2.4 Rumelhart (1967) .......................................... 9
1.2.5 Cepeda et al. ............................................. 10
1.2.5.1 Experiment 1 ........................................... 11
1.2.5.2 Experiment 2 ........................................... 12
1.2.5.3 General Discussion of Cepeda et al. Results .... 14
1.3 Leading Model Theories ........................................ 15
1.3.1 Encoding variability ...................................... 15
1.3.1.1 Challenges to Encoding Variability ............... 16
  – Ross & Landauer (1978)
1.3.2 Inattention theory ....................................... 20

2 Raaijmakers model
2.1 Search of Associative Memory (SAM) ....................... 21
2.2 Modifications to SAM ......................................... 22
2.2.1 Contextual Fluctuation Theory ........................... 22
2.2.2 Short-term memory store ................................... 24
2.3 Capabilities of Raaijmakers (2003) model ............... 25
2.3.1 Fitting Glenberg (1976) Experiment 1 .................. 25
2.3.2 Fitting Young (1971) data ............................... 27
2.3.3 Fitting Rumelhart (1967) data ........................... 27
2.3.4 Accounting for Ross & Landauer (1978) ............. 27
2.4 General critique of Raaijmakers (2003) model ........ 29
2.5 Meta analysis of Raaijmakers (2003) ........................ 30
3 Pavlik & Anderson model
3.1 Adaptive Character of Thought-Rational (ACT-R) …… 32
3.2 Modifications to ACT-R …………………………………… 32
3.3 Capabilities of the Pavlik & Anderson (2005) model … 33
3.3.1 Fitting Pavlik & Anderson (2005) experiment ……… 34
3.3.2 Fitting Glenberg (1976) Experiment 1 ………………… 34
3.3.3 Fitting Young (1971) experiment …………………… 34
3.3.4 Fitting Rumelhart (1967) experiment ………………… 34
2.4 General critique of Pavlik & Anderson (2005) model … 38
2.5 Meta analysis of Pavlik & Anderson (2005) model …… 38

4 The Extended TAC-V model
4.1 The original TAC-V model ……………………………… 42
4.1.1 Finding Optimal ISIs with TAC-V model ………… 46
4.2 Efforts taken to improve the TAC-V model ………… 48
4.2.1 Variant 1 of TAC-V:
Incorporating fixed encoding probability ……………… 49
4.2.2 Variant 2 of TAC-V:
Incorporating variable difficulty of item encoding …… 50
4.2.3 Variant 3 of TAC-V:
Incorporating conditioning on recall at S2 ………… 51
4.2.4 Addressing how to incorporate difficulty ………… 53
4.2.4.1 Item difficulty via encoding probability ………… 53
4.2.4.1.1 Fitting Cepeda et al. data ……………… 53
4.2.4.1.2 Problems with this approach ………… 56
4.2.4.2 Item difficulty via modulation of
contextual distance parameters ……………… 56
4.2.4.2.1 Fitting Cepeda et al. data ……………… 56
4.2.4.2.2 Fitting Glenberg data ……………… 58

5 Concluding Remarks ……………………………………… 60

6 References ………………………………………………… 61
1 Introduction

1.1 The Spacing Effect – Overview

The main observation of the spacing effect is that as spacing between opportunities to study information increases, long-term retention of that information improves.

To observe spacing effects in a typical paired-associate learning task, one may give a subject two study opportunities, $S_1$ and $S_2$, to learn the paired-associate $A - B$. These study opportunities can either involve presentation–study sessions of the $A - B$ pair, where the complete pair is presented to the subject, or test–study sessions where the subject is given $A$, and asked to produce $B$; following this testing, the correct response $B$, is presented to the subject. In this style of experiment, the time elapsed between $S_1$ and $S_2$ is denoted as the inter-study interval (ISI).

Following $S_2$, an evaluation $T$ is administered (e.g., for each $A - B$, given component $A$, recall component $B$) to the subject to determine information retention, where the time elapsed between $S_2$ and $T$ is denoted as the retention interval (RI). The structure of this experiment is represented as such:

$$S_1 \xrightarrow{ISI} S_2 \xrightarrow{RI} T$$

The remainder of CHAPTER 1 will be devoted to descriptions of influential experiments on the spacing effect and their findings. In addition, two major theories related to modeling the spacing effect, encoding variability and inattention theory, will be discussed.

1.2 The Spacing Effect in Experimental Literature

In the past century, many experiments have been attempted in order to demonstrate the spacing effect in learning and memory. These efforts have produced valuable insights, which have since become crucial to the theoretical analysis of the spacing effect. In addition, the data collected by these studies has provided a gold standard in efforts to model the spacing effect.

1.2.1 Glenberg (1976)

In 1976, Arthur Glenberg conducted an experimental study on the spacing effect, centered on a paired associate learning task of the same structure given in SECTION 1.1. In Experiment 1 of Glenberg (1976), subjects were given two study opportunities $S_1$ and $S_2$ to learn a series of pairs of common four-letter nouns $A - B$, consisting of a
cue noun $A$, and a target noun $B$. Following these presentations, an evaluation $T$ would be administered in which subjects would be given the cue noun $A$, and asked to recall the target noun $B$.

The objective of this experiment was to determine the effect of both inter-study interval and retention interval on recall. Values of retention interval studied were 2, 8, 32, and 64 days. For each retention interval studied, the experiment looked at corresponding inter-study interval values of 1, 4, 8, 20, 40 days, as well as an inter-study interval of negligible length (25 minutes).

The results of Experiment 1 in the Glenberg (1976) paper are given in FIGURE 1.1. For a retention interval of 2 days, a clear non-monotonic relationship can be observed for inter-study interval, with an inter-study maximizing probability of recall (denoted as the optimal inter-study interval) appearing at 4 days. Although this non-monotonic relationship is still apparent for the 8-day retention interval, with an optimal inter-study interval appearing at 8 days, the graphical peak is not as pronounced as with the 2-day case. At retention intervals of 32 and 64 days, a monotonic relationship is observed with inter-study interval. Based on the results of Glenberg (1976) Experiment 1, it is clear that optimal inter-study intervals exist at retention intervals of about a week or less. However, with retention intervals larger than this, these experimental results do not indicate such optimal inter-study intervals exist.

![FIGURE 1.1](Graph of Glenberg (1976), Experiment 1 results.)
1.2.2 Young (1971)

In the Young (1971) publication, a paired associate learning experiment was conducted, in which subjects would learn consonant trigrams, paired with a corresponding digit ranging from 0-9.

In this experiment, a learner would be given two presentation-study sessions $S_1$ and $S_2$ for $A - B$ pairs, consisting of a cue string (the consonant trigram) $A$, and a corresponding digit to be recalled $B$. Following these presentations, an evaluation $T$ would be administered in which the subject would be given the cue string $A$, and asked to recall the corresponding digit $B$.

The Young (1971) experiment used a retention interval of 10 days. Corresponding inter-study intervals of 1 to 17 days, as well as an inter-study interval of negligible length (30 minutes), were also used.

The results of this experiment in the Young (1971) paper are graphically represented in FIGURE 1.2. By visual inspection, an optimal inter-study interval occurs at 7 days. The findings of this particular experiment were among the first to demonstrate a non-monotonic relationship between optimal inter-study interval and retention interval. In other words, that a finite inter-study interval can exist with which probability of successful recall of information is maximized.

FIGURE 1.2
Graph of Young (1971) results.
1.2.3 Pavlik & Anderson (2005)

In Pavlik and Anderson (2005), a foreign language learning exercise was conducted using Japanese-English word pairs. In this exercise, subjects started with a study session, in which they had to learn 104 Japanese-English word pairs over the course of 12 study blocks, each containing 40 study opportunities. These study opportunities consisted of either a presentation-study of the Japanese-English pair, where the complete pair would be presented to the subject, or a test-study session, in which the subject would be given a Japanese word, and asked to produce the corresponding English word; following this testing, the correct English word would be presented to the subject.

Each of the Japanese-English word pairs was tested 1, 2, 4, or 8 times, with spacings of 2, 14, or 98 intervening study opportunities. Although this design results in 12 conditions to compare, the 8 test / 98 spacing condition was disregarded due to time constraints. Therefore, 11 conditions were considered for this experiment, in which 8 word pairs were randomly selected from the 104-word pair pool and used in each condition. For each subject, only responses on the chosen word pairs were considered for analysis (8 for each condition, 88 total). The remaining 16 word pairs were used as filler items and left unanalyzed.

Following a retention interval of 1 or 7 days, subjects participated in an evaluation to determine retention of knowledge. For this, subjects were tested on each of their 88 selected Japanese-English word pairs 4 times with 98 intervening tests. The format of these tests was identical to the test-study opportunities given in the study session.

**FIGURE 1.3**

Graph of learning session results for Pavlik and Anderson (2005).
FIGURE 1.4
Crossover interactions in Pavlik and Anderson (2005) results.

FIGURE 1.5
Graph of evaluation results for Pavlik and Anderson (2005).
FIGURE 1.3 gives average recall accuracy obtained from subjects within the study session as a function of times studied and intervening study opportunities. In the study session, as spacing decreased between word pairs, better performance resulted due to limited forgetting. In both the 2 and 14 spacing categories, the probability of answering a word-pair correctly came close to 100 percent within 5 trials.

FIGURE 1.4 provides average probability of correctness for the 2, 14, and 98 spacing categories, both at the end of the study session and the first test of the evaluation. For all spacing conditions, forgetting naturally occurred. In addition, a spacing effect was obtained, where the probability of recall dropped the least for the 98 spacing condition. This finding suggests that wider spacing results in better long-term performance. Moreover, this figure suggests that spacing affects the forgetting rate, since forgetting was larger for the shorter spacing conditions.

FIGURE 1.5 details average performance in the testing session on word-pairs from each of the 11 study session categories. Performance on the first trial varied, ranging from over 50 percent probability of correctness for the 8 test/14 spacing and 4 test/98 spacing categories, to about 10 percent for the 2 test/2 spacing category. However, by the fourth trial, word pairs from each of the 11 $S_1$ categories had somewhere between a 70 to 90 percent chance of being answered correctly. Interestingly, performance was best for the longer spacing conditions not only on the first trial of the evaluation session, but on all subsequent trials until asymptotic performance within the session was achieved.
1.2.4 Rumelhart (1967)

In Rumelhart (1967), a spacing effect based experiment was conducted which involved teaching subjects a list of 66 paired associates, structured as consonant-vowel-consonant trigrams, paired with one digit, either a 3, 5, or 7. The structure of this experiment involved one test-based learning session, in which subjects were tested on each paired associate 6 times at inter-test intervals ranging from 1 to 10 pairs. In selecting the appropriate lag sequences to test, the sequences chosen are given in FIGURE 1.6, as well as their corresponding results.

![FIGURE 1.6](image)

In these findings, forgetting can be observed, which increases with inter-study interval, in addition to improved memory through spacing, which may also increase with inter-study interval. In situations where short spacings are used, followed by testing after a long spacing, drops in performance occur, as with the lag sequence 1-1-10-10-10. However, when long spacings are used, followed by testing after a long spacing, no drop in performance occurs, as with the lag sequence 10-10-10-10-10. Because a retention interval of 10 is reflected in the final data point for each lag sequence, it becomes worthwhile to compare this point across lag sequences. Regardless of the lag sequence, an accuracy of about 90 percent can be observed for the final data point. In situations where the spacing of the next to last data point was 6 or 10, the final data point would constitute the peak of the lag sequence. However, when the spacing of the second to last data point was 1 or 3, this data point would instead constitute the peak, with accuracies ranging from 90 to 95 percent in the 8 example sequences given.
Cepeda et al. (2006) published the findings of two spacing effect based experiments. The purpose of these two experiments was to provide experimental data based on two different retention intervals: one of 10 days, and one of 6 months.

The retention interval of 10 days, used in the first experiment (Experiment 1), had been used in four earlier spacing effect experiments, each of which displayed some type of non-monotonic relationship between retention interval and inter-study interval. The results of these four studies (Ausubel (1966); Childers & Tomasello (2002); Edwards (1917); Glenberg & Lehmann (1980)) can be viewed in FIGURE 1.7. As such, the primary purpose of the first experiment was to replicate the non-monotonic relationship displayed by the results of these four studies.

By contrast, the retention interval of 6 months used in the second experiment (Experiment 2) was believed by Cepeda et al. to be much longer than that used by any prior spacing effect study. As such, by conducting an experiment using a retention interval of this length, it was believed that theoretical analysis of the spacing effect could be expanded.

FIGURE 1.7
Results of Ausubel (1966); Childers & Tomasello (2002); Edwards (1917); Glenberg & Lehmann (1980).
Figure from Cepeda et al. (2006) paper.
1.2.5.1 Experiment 1

For the first experiment of Cepeda et al. (2006), henceforth denoted as Experiment 1, a foreign language learning exercise was conducted, using Swahili–English word pairs. In this experiment, subjects were presented 40 Swahili–English word pairs individually in a pre–experiment learning session. Immediately following this learning session, subjects proceeded to a test–study session, in which the Swahili component of each word pair was presented, and subjects were asked to produce the English component. In the event of producing an incorrect answer, subjects were provided an opportunity to restudy the Swahili–English word pair before proceeding to the next question. Likewise, in the event of producing a correct answer, subjects immediately proceeded to the next question. For each word pair incorrectly answered, subjects were re-tested on it some random number of tests later until a correct answer was produced. This process continued until all word pairs had been answered correctly.

Following the test–study session, subjects participated in a second test–study session of the same structure as the first one, following an inter-study interval of 1, 2, 4, 7, or 14 days, or one of negligible length (5 minutes).

After the second test-study session, knowledge retention was tested in an evaluation, following a retention interval of 10 days for all subjects. For this evaluation, subjects were presented with the Swahili component of each word pair and asked to produce the English component, as in the two test-study sessions. However, the opportunity to re-study a pair answered incorrectly was not given, nor was the opportunity to be re-tested on it.

FIGURE 1.8
Results of Cepeda et al. (2006), Experiment 1. Figure from Cepeda et al. (2006) paper.
The results for Experiment 1 are graphically represented in FIGURE 1.8. In the second test-study session, for the initial tests on each Swahili–English word pair, and prior to any re-testing of incorrectly answered questions, a traditional forgetting function can be observed. In the final testing session, a clear non-monotonic curve can be observed, with an optimal spacing interval of 1 day. This non-monotonic relationship between retention interval and spacing interval is consistent with the findings of the four prior spacing effect studies cited by Cepeda et al. (2006), using retention intervals of similar length (1-2 weeks) Ausubel (1966); Childers & Tomasello (2002); Edwards (1917); Glenberg & Lehmann (1980). However, based on the findings of Experiment 1 alone, it remains unclear whether this non-monotonic relationship continues to hold for retention intervals longer than 2 weeks.

1.2.5.2 Experiment 2

Although many experiments have been conducted to demonstrate the spacing effect in learning and memory, virtually none have studied retention intervals longer than 2 weeks. This lack of experimental history poses problems for the spacing effect going beyond theoretical analysis. For instance, educators typically want to determine how to distribute classroom lessons to maximize retention of information for 6 months or more, as opposed to 2 weeks or less. Because of a lack of experimental findings based on long retention intervals (greater than 2 weeks long), practical applications of the spacing effect such as this one have not been widely considered.

For Experiment 2, the objective was to measure optimal spacing for retention intervals 6 months in length – something that, according to Cepeda et al., had not been attempted prior. Based on preliminary findings, it was determined that testing acquisition foreign language pairs (as in Experiment 1) would not produce useful results at long retention intervals. Instead, it was decided that subjects would be tested on two alternate forms of associative pair based information: learning obscure facts and obscure image recognition.

Structurally, Experiment 2 was organized in much the same way as Experiment 1. However, to test with two distinct forms of associative pair based information, Experiment 2 was organized into 2 sections: section 2a for the obscure fact learning, and section 2b for obscure image recognition. As such, for each step of the experiment, an Experiment 2a variant was followed by a corresponding Experiment 2b variant. This described experiment structure is visually represented in FIGURE 1.9.

In Experiment 2, obscure fact learning involved learning the answer component of a question-answer pair for a given obscure fact. For example, for the question component “Who first synthesized the chemical compound borazine in 1926?” the corresponding answer component would have been “Alfred Stock.” Likewise, obscure image recognition involved learning the answer component of a image-answer pair for an image of an obscure object, where the image component consisted of a photograph of the object and an associated description (1-2 sentences) of it. For example, given a photograph of a
1965 Aston Martin DB5, and the associated fact “Name this vehicle, driven by James Bond in the 1964 movie Goldfinger,” the subject would have been expected to produce the name of the object in the photograph. In both the fact learning and image recognition sections, partial answers containing distinctive information from the complete answer would have been recognized as correct (e.g., answering “Aston Martin DB5” instead of “1965 Aston Martin DB5”).

In Experiment 2, subjects started with an initial test-study session $S_1$, in which they were presented with a question component, either a fact-based question for section 2a, or a photograph with an associated description for section 2b, and asked to produce an answer component for each of 23 individual items in both sections. Regardless of answer correctness, subjects were given the opportunity to briefly re-study the question-answer pair before proceeding to the next question. In addition, unlike in Experiment 1, subjects were not re-tested on incorrectly answered questions.

Following session $S_1$, subjects participated in a second session $S_2$ of the same structure as $S_1$, following an inter-study interval of 1, 7, 28, 84, or 168 days, or one of negligible length (5 minutes).

After the second session $S_2$, knowledge retention was tested in an evaluation session $T$, following a retention interval of 168 days for all subjects. For the final evaluation session $T$, subjects would be presented with the question or image component of each paired associate and asked to produce the answer component, as in study sessions $S_1$ and $S_2$. However, the opportunity to re-study a pair answered incorrectly would not be given, nor would the opportunity to be re-tested on it.

**FIGURE 1.9**
Structure of Cepeda et al. (2006) Experiment 2

$$
S(2a)_1 \rightarrow S(2b)_1 \xrightarrow{ISI} S(2a)_2 \rightarrow S(2b)_2 \xrightarrow{RI} T(2a) \rightarrow T(2b)
$$

$ISI = 25\text{ min}, 1, 7, 28, 84, \text{ or } 168\text{ days}$

$RI = \text{FIXED at } 168\text{ days}$
A graph of the results for sections 2a and 2b of Experiment 2 is given in FIGURE 1.10. In the second test-study session (denoted as $S_2$ in FIGURE 1.9) for sections 2a and 2b, traditional forgetting functions can be observed. In the final testing session (denoted as $T$ in FIGURE 1.9), clear non-monotonic curves can be observed for both sections. This non-monotonic relationship between spacing interval and final retention is consistent with the both findings of Experiment 1, and those of the four spacing effect studies cited in Cepeda et al. (2006) (Ausubel (1966); Childers & Tomasello (2002); Edwards (1917); Glenberg & Lehmann (1980)).

For both sections 2a and 2b of Experiment 2, optimal spacings occurred at 28 days. This is an interesting observation, considering the difference in learning material given to participants in these 2 sections. For this reason, it may be possible that optimal spacing is not dependent on the type of learning material being studied.

### 1.2.5.3 General Discussion of Cepeda et al. Results

For Experiments 1 and 2 (parts 2a and 2b) of Cepeda et al., the ratios of optimal ISI to RI are presented in the above graph. In Experiment 1, for an RI of 10 days, an optimal ISI of 1 day was observed. Likewise, in Experiment 2, both parts 2a and 2b, for an RI of 168 days, an optimal ISI of 28 days was observed.
Despite the different lengths of retention interval for Experiments 1 and 2, an important conclusion can be reached about both experiments. In both cases, a non-monotonic relationship can be observed between inter-study interval and retention interval. In addition, in comparing the optimal inter-study interval to retention interval ratios of Experiments 1, 2a, and 2b, it is clear they fall within a similar range. However, for a fixed optimal ratio to occur between optimal ISI and RI, a linear relationship of the form $Y = ax$ must exist, where the slope $a$ represents the optimal ratio.

To determine whether this linear relationship exists between optimal inter-study interval and retention interval, a log-log equation conversion is employed. This results in a transformation of equations of the form $Y = ax^b$ to a linear equation $\log(y) = b \log(x) + \log(a)$. Based on this relationship, equations with linear relationships (e.g., $b = 1$) should result in a log-log equation with a slope of 1.

Setting $Y = RI$ and $x = \text{optimal ISI}$, for all data collected for Experiments 1 and 2, a log-log equation conversion results in a linear equation with slope $b = 1.1811$, and $y$-intercept $\log(a) = -2.7195$. Since slope equals a value other than 1, a linear relationship does not exist between optimal ISI and RI for Experiments 1 and 2. This finding calls into question whether a special ratio exists between optimal inter-study interval and retention interval.

1.3 Leading Model Theories

Two classes of theories have been widely proposed to explain the spacing effect. These classes are known as encoding variability and inattention theories.

1.3.1 Encoding variability

According to encoding variability theory, memory is stored in specialized traces, containing details explicit to the memory itself, as well as information derived from the learning context. This context contains implicit features from the learning environment, such as the color of the surrounding walls, temperature of the room, and so forth. This context is believed to change over time based on random fluctuation of environment features, known as a random walk. At any given instant in time, context can move toward, or away from, the context of a previous memory encoding. The range of possible locations the context could wander to, known as variance, grows linearly with time.

Recalling the experiment structure from SECTION 1.1, given two study sessions $S_1$ and $S_2$, and a final testing session $T$, each will have a unique context encoding. These context encodings can be denoted as the quantities $C_{S_1}, C_{S_2}$, and $C_T$, respectively. **FIGURE 1.11** illustrates a one-dimensional random walk between these contexts.
Retrieval at the evaluation session $T$ depends on the similarity of context $C_{T}$ to the contexts for study session $S_1$ or $S_2$. The closer the testing context $C_{T}$ moves to one of these study session contexts, the more likely it becomes for memory recall to occur. Since context fluctuates as a function of time, it is believed that encoding variability theory is consistent with the interaction between retention interval and inter-study interval.

**FIGURE 1.11**

Example of one-dimensional contextual random walk. Figure from Cepeda et al. (2006) paper.

1.3.1.1 Challenges to Encoding Variability – Ross & Landauer (1978)

Ross and Landauer (1978) presents a major challenge to encoding variability theory. They argued that in encoding variability theories, the traces for the two study sessions are laid down independently of each other. As a result, the probability of recall at final testing should simply be the probability that *either* of the two traces is recalled. In quantitative terms, if $P(T \mid S_1)$ is the probability of recall at testing given the trace laid at $S_1$, where $S_i$ equals either $S_1$ or $S_2$, then the probability of recall at evaluation given both of the traces should be $P(T \mid S_1, S_2) = 1 - (1 - P(T \mid S_1))(1 - P(T \mid S_2))$. 

To demonstrate that this relationship does not occur in learning and memory, Ross and Landauer (1978) presents the findings of two experiments, in which items presented twice and tested once show a spacing effect, but pairs of two items presented and tested once do not.

For Experiment 1, recognition testing of a series of 300 rare English words was conducted. Initially, subjects were presented each of the 300 words in an initial learning session. For each word presentation given, subjects were presented 2 words from the 300-word pool. These presentation instances were structured such that each word from the body of 300 was presented either once or twice in the study session. Out of the 300 words used, 180 were intended for analysis, and were divided into 5 groups of 36, with their presentations (between two of the same item, or two distinct items) spaced at 0, 1, 3, 9, or 27 intervening presentations. The remaining 120 words were used as buffer items and not considered for analysis, or incorporated into later parts of the experiment.

Following the initial learning session, subjects were put through a testing session, in which a series of 150 recognition tests were given, each bearing 2 different words. In each recognition test, subjects were asked whether one, both, or neither of the words were previously studied. For each of the 5 spacing categories from the presentation session, 24 cards were used. Of these, 12 of the cards would include two words from the spacing category, 8 of the cards would include 1 word from the spacing category, and 1 word from another spacing category, and 4 would include 1 word from the spacing category paired with a word not appearing in the original 300-word list. These 150 recognition tests appeared in no special order for subjects.

In recognition sessions where two words from the original 300-word list were presented, the probability of recognizing at least one of the words was denoted as the OR score. According to Ross and Landauer (1978), this is critical in determining whether spacing effects occur in pairs of two words presented and tested once. The OR scores for all five spacings, for pairs of two words presented once at different spacing intervals, and tested together, is shown by the darkened circles in FIGURE 1.12. Likewise, the OR scores for all five spacings, for pairs of two words each presented once at different spacing intervals, and tested separately, is shown by the white circles in FIGURE 1.12. Although neither of these categories demonstrates an appreciable spacing effect, for single words presented twice in the presentation-study session, a spacing effect can be observed. This effect is denoted by the filled in squares in FIGURE 1.12.

In previous spacing effect experiments, it was found that spacing effects were not as pronounced on recognition tasks (such as those in Glenberg (1976), Experiment 3; Underwood, Kapelak & Malmi (1976), Experiments 1 and 2), as in free recall tasks (as in Glenberg (1976), Experiment 1). Therefore, a second experiment was conducted with similar structure to Experiment 1, but with an incorporated a free recall task between the presentation-study session and the evaluation session. In the free recall task, subjects
were given blank sheets of numbered paper, and told to write down every word remembered from the presentation-study session. Following this test, an evaluation session was conducted of identical structure to Experiment 1.

**FIGURE 1.12**

*Figure 1 from Ross & Landauer (1978)*

**FIGURE 1.14** presents the findings of the free recall part of Experiment 2. Although a spacing effect can be observed for single words presented twice in the presentation-study session, this effect does not seem to occur for pairs of two words each presented once. **FIGURE 1.15** presents the findings of the recognition section following the free recall task. As in Experiment 1, though a spacing effect can be observed for single words presented twice in the presentation-study session, no significant spacing effect seems to occur for pairs of 2 words each presented once.

Since the findings of both Experiments 1 and 2 seem to contradict encoding variability theory (based on the mathematical logic given earlier), they call into question the validity of all encoding variability based models. However, since these findings were published, two encoding variability based models have been created which seem to account for them, Raaijmakers (2003) (discussed in **CHAPTER 2**) and Temporal Associative Context Variability (TAC-V) (discussed in **CHAPTER 4**).
FIGURE 1.13
Figure 3 from Ross & Landauer (1978)

Fig. 3. Recall performance in Experiment 2. Data are proportion of words recalled for words presented twice in study list (one-word-twice) and OR scores for sets of two words each presented once (two-words-once), as a function of spacing between presentations.*

FIGURE 1.14
Figure 2 from Ross & Landauer (1978)

Fig. 2. Recognition performance in Experiment 2. Data are proportion correct for words presented twice in the study list (one-word-twice) and OR score for sets of two words each presented once (two-words-once), as a function of spacing between presentations. Experiment 2 was identical to Experiment 1 except for the use of less common words and a different retention interval task.³
1.3.2 Inattention

Inattention theory states that, given two study sessions $S_1$ and $S_2$ to learn a given item, failure to attend to session $S_2$ will occur if spacing between $S_1$ and $S_2$ is too short. When failure to attend to Session $S_2$ occurs, it can’t contribute to long-term memory of the given items. This means the information introduced in $S_2$ can’t form a new memory trace, or strengthen the trace formed at $S_1$.

One way to incorporate inattention theory into a probabilistic model involves use of a short-term memory store, as in Mensink and Raaijmakers (1989). When a new item $i$ is presented at study session $S_1$, it will enter the short-term memory store and be written as a new memory image to long-term memory. Following this long-term memory encoding, when the same item is presented at study session $S_2$, two possible things may happen. If the item $i$ is still in the short-term memory store at study session $S_2$, no additional information will be added to it. Therefore, no information from study session $S_2$ will be added to long-term memory. Likewise, if item $i$ is not present in short-term memory during study session $S_2$, it will either be recorded as a new memory image, or be incorporated into the memory image formed during $S_1$. As such, if the spacing interval between $S_1$ and $S_2$ is not long enough for the $S_1$ presentation of item $i$ to leave short-term memory, no information pertaining to $S_2$ will be attended to.

Based on this mechanism of memory encoding, the optimal spacing interval becomes dependent on properties imposed by the short-term memory store. Therefore, the possibility of obtaining the optimal spacing interval as a function of retention interval is lost.
2 Raaijmakers model

Raaijmakers (2003) developed an extension to the Search of Associative Memory (SAM) theory of recall, first mentioned in Raaijmakers and Shiffrin (1980).

2.1 Search of Associative Memory (SAM)

Originally, SAM was developed as a probabilistic model of recall from memory, capable of explaining many observable results in both free and cued recall. According to SAM theory, memory is represented by specialized “images,” containing details explicit to the memory itself. In addition to this, secondary explicit information bearing a relevant association to the memory is stored as associative contextual information (or “cues” for the memory image). Furthermore, in addition to memory-explicit information, implicit information related to the context of the environment during learning is stored as implicit contextual information. As an example of the storing of a SAM-based memory image, consider an English-speaker learning the French word *dirigeable* (which translates into English as ‘airship’). In the resulting memory image, the associative recall cue would be the French word *dirigeable*, and the direct memory would be of the corresponding English word. In addition, information not directly related to the learning task (such as the temperature of the classroom and color of the walls) would be integrated into the memory image as implicit contextual information.

In SAM, retrieval depends on the level of associative strength a set of recall cues (or components of a recall cue) to a given memory. To determine the overall associative strength $A(i, Q_1...Q_n)$ of a set of recall cues $\{Q_1...Q_n\}$ to a memory image $i$, the following equation is used:

$$A(i, Q_1...Q_n) = \prod_{j=1}^{n} S(i, Q_j) \quad \text{EQUATION 2-1}$$

In this, $S(i, Q_j)$ represents the associative strength of one cue component $Q_j$ to a memory image $i$. One key property of the equation above is that all presented cues must have some association to the memory image (e.g.: $S(i, Q_j) > 0$). As such, memory traces having no association with one or more recall cues will have a net associative strength of zero.

For a given set of cues $\{Q_1...Q_n\}$, an associative strength $A(k, Q_1...Q_n)$ will exist to every memory image $k$ in long-term store. The summation of these associative strengths $\sum A(k, Q_1...Q_n)$ represents the associative strength between cues $\{Q_1...Q_n\}$ and all memory images in long-term store. With this, the probability of accessing memory image $i$ with associative strength $A(i, Q_1...Q_n)$, as opposed to any other memory image in long-term store, becomes:
\[ P_{Access}(i) = \frac{A(i, Q_1 \ldots Q_n)}{\sum\limits_{k} A(k, Q_1 \ldots Q_n)} \]  \hspace{1cm} \text{EQUATION 2-2}

Once memory image \( M_i \) is successfully accessed, it becomes possible to recall relevant information from it. However, due to inevitable noise in memory images, successful recall of encoded information is not guaranteed. Going back to the French-English language example, if the presented cue ‘airship’ results in successful sampling of the relevant memory image, the probability of recalling the French word \emph{dirigeable}, conditional on having accessed memory image \( i \), becomes:

\[ P_{Recall}(i) = 1 - \exp\left[-\sum_{j=1}^{n} S(Q_j, i)\right] \]  \hspace{1cm} \text{EQUATION 2-3}

Where \( P_{Recall}(i) \) is given by an exponential function of the sum the associative strengths of each cue component \( Q_j \) to the memory image \( i \). Should this retrieval attempt fail, it is repeated a maximum of \( L_{\text{MAX}} \) times before failing altogether. Therefore, the probability of recalling information from a memory image \( i \), conditional on having accessed this memory image, is represented by the probability of the item being sampled at least once, times the probability of a successful recall:

\[ P_{Total}(i) = [1 - (1 - P_{Access}(i))^{L_{\text{MAX}}}] P_{Recall}(i) \]  \hspace{1cm} \text{EQUATION 2-4}

2.2 Modifications to SAM

Despite successes of the SAM model demonstrated by Raaijmakers and Shiffrin (1980), it originally lacked the capability account for forgetting (decay) of information from memory. As such, it was incapable of accounting for spacing effects in distributed practice. To remedy this, an extension of SAM was implemented, known as the contextual fluctuation theory, by Mensink and Raaijmakers (1988, 1989). Building on this extension, support for a short-term memory buffer was included by Raaijmakers (2003).

2.2.1 Contextual Fluctuation Theory

According to contextual fluctuation theory, environmental context consists of discrete elements, which are binary valued – either present or absent – and fluctuate randomly between these two states. These elements consist of features such as the color of the surrounding walls, temperature of the room, and so forth.

When all of these discrete elements are combined, they form a \emph{binary feature vector}, capable of defining environmental context at any given instant. Since each discrete element in the binary feature vector is in random fluctuation, environment context has the
capability of moving toward previous states – increasing the probability of recalling memories previously encoded. In addition, environment context has the ability to move away from its current configuration – resulting in gradual forgetting of newly encoded memories.

According to contextual fluctuation theory, the probability of sampling an image $i$ from memory depends on the temporal lag (the retention interval) between initial storage and the time of sampling, denoted $t$, denoted by EQUATION 2-5.

$$P_{Sampling}(i,t) = \frac{c(i,t)I(Q_1\ldots Q_n,i,t)}{c(i,t)I(Q_1\ldots Q_n,i,t) + \sum Z(Q_1\ldots Q_n,k)}$$

EQUATION 2-5

In this equation, the value $c(i,t)$ represents the implicit contextual strength between the environment and memory image $i$ after a retention interval of $t$ seconds, and $I(Q_1\ldots Q_n,i,t)$ represents the associative strength between a series of presented cues $\{Q_1\ldots Q_n\}$ and memory image $i$ after $t$ seconds. In the denominator, $\sum Z(Q_1\ldots Q_n,k)$ represents an interference factor involving associative strengths between a series of cues $\{Q_1\ldots Q_n\}$ and all other possible memory images $k$.

Given a successful sampling of memory image $i$ after retention interval of $t$ seconds, the probability of successfully recalling information from the image becomes EQUATION 2-6.

$$P_{Recall}(i,t) = 1 - \exp[-\theta c(i,t) + I(Q_1\ldots Q_n,i,t))]$$

EQUATION 2-6

In the equation above, where $\theta$ is incorporated to distinguish whether presentation-study or test-study sessions were used during the second learning session $S_2$, prior to evaluation $T$. Based on general assumption, it is stated that probability of successful recall should be higher if a presentation-study took place (where a paired associate $A \rightarrow B$ was presented and implicitly learned) than if a test-study session occurred (with only the component $A$ being presented, and no guaranteed explicit learning). Through this, separate values for $\theta$ are assumed for these 2 cases, with $\theta = 1$ if presentation-study sessions were used and $0 \leq \theta \leq 1$ if test-study sessions were used.

The Raaijmakers (2003) model defines the implicit contextual strength between the environment and memory image $i$ after a retention interval of $t$ seconds $c(i,t)$ as the number of contextual elements present at the time of retrieval that were also present at the initial presentation, defined in EQUATION 2-8 below.

$$c(i,t) = c(0)e^{-at} + Ks(1 - e^{-at})$$

EQUATION 2-8
In this equation, \( c(0) \) denotes the number of elements active for memory at time \( t = 0 \), and \( K \) represents the total number of active and non-active contextual elements. The parameters \( s \) and \( \alpha \) are tied to the rate of contextual element fluctuations between active and non-active states. Building on **EQUATION 2-8**, the number of available contextual elements from two study opportunities, given an inter-study interval of \( t_1 \) seconds and a retention interval of \( t_2 \) seconds, is denoted below.

\[
c_2(t_1, t_2) = c_2(t_1, 0)e^{-\alpha t_2} + K_2(t_1)s(1 - e^{-\alpha t_2})
\]

**EQUATION 2-9**

In this equation, \( K_2(t_1) \) represents the total contextual elements stored during the two study opportunities with spacing \( t_1 \) seconds. This results in a dependency on \( t_1 \), since as this value decreases, the contexts at the two study opportunities become more similar, resulting in fewer stored contextual elements. Aside from this dependency, the equation is essentially equivalent to the former one presented. Based on this, the relationship \( c_2(t_1, 0) = c(0) \) follows. In addition, since \( K_2(t_1) \) is denoted as the sum of the elements active at one of the two study opportunities minus the elements active at both study opportunities, as in **EQUATION 2-10**.

\[
K_2(t_1) = 2c(0) - c(t_1)
\]

**EQUATION 2-10**

In this extended SAM model, the total probability of information recall from memory image \( i \) is the same as in the original SAM model, with the exception of a conditional dependency on a retention interval of \( t \) seconds, as in **EQUATION 2-11**.

\[
P_{\text{Total}}(i, t) = [1 - (1 - P_{\text{Sampling}}(i, t))^{L_{\text{MAX}}}]P_{\text{Recall}}(i, t)
\]

**EQUATION 2-11**

### 2.2.2 Short-term memory store

After the first study \( S_1 \) of a given item, it enters the short-term memory with probability \( p \), forming a new memory image in long-term memory. After the second study \( S_2 \) of the item, three possibilities exist:

1. The item studied at \( S_1 \) is still in the short-term memory, resulting in no additional information being stored in long-term memory.
2. The item studied at \( S_1 \) is in long-term memory and is retrievable, resulting in information being added to the original memory image.
3. The item studied at \( S_1 \) is in long-term memory yet is irretrievable, resulting in the formation of a new memory image in long-term memory.

The probability that an item remains in short term memory buffer after \( t \) seconds is given as \( P_{\text{STS}}(t) = e^{-\lambda t} \), where \( \lambda \) represents a decay constant.
These modifications to SAM, combined with those presented through introduction of contextual fluctuation (see SECTION 2.2.1) constitute the major changes between the original SAM model and the Raaijmakers (2003) model.

2.3  Capabilities of Raaijmakers (2003) model

The Raaijmakers (2003) publication detailed efforts to model experimental data from previous spacing effect experiments. Namely, the data from Experiment 1 in the Glenberg (1976) paper introduced in SECTION 1.2.1, the experimental data from Young (1971) introduced in SECTION 1.2.2, and Rumelhart (1967) experiment introduced in SECTION 1.2.4.

FIGURE 2-1 provides a description of each of the parameters in the Raaijmakers (2003) model, as well as values used for fitting data from these 3 experiments. To obtain these parameter fits, a general-purpose optimization program (known as MINUIT) was employed. This program worked by scanning over all the Raaijmakers (2003) model parameters, given in FIGURE 2-1, and adjusting them to minimize the root mean squared error for a set of experimental data with \( n \) points, defined in EQUATION 2-12.

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} (obs_i - exp_i)^2}{n}} \tag{2-12}
\]

In the \( RMSE \) definition, \( n \) is the number of data points and the values for \( obs_i \) and \( exp_i \) are the observed and predicted probabilities of successful recall for item \( i \) out of \( n \) items. The summation in the equation indicates a sum over all data points, and the division by \( n \) gives the mean of the squared errors for the \( n \) items.

2.3.1  Fitting data from Experiment 1, Glenberg (1976)

FIGURE 2-2 details the results of the Raaijmakers (2003) model’s attempt to fit the data from Glenberg (1976), Experiment 1. Parameters obtained for this fit are given in figure FIGURE 2-1, which resulted in a \( RMSE \) value of 1.32. This \( RMSE \) value is not as good as those for the Young (1971) and Rumelhart (1967) experiments to be discussed shortly. In general, the Raaijmakers (2003) model does capture a general non-monotonic trend present in the Glenberg (1976) Experiment 1 data. However, in the 6 and 24-day spacing categories, the Raaijmakers (2003) model does not capture the degree of amplitude in the peaks of this experimental data. In addition, for the 96 and 192-day categories, a non-monotonic trend was captured which was not apparent in the Glenberg (1976) experimental data.

Table 1
Parameters and their values in the fits to various data sets

<table>
<thead>
<tr>
<th>Parameters and their meaning</th>
<th>Runelhart data</th>
<th>Young data</th>
<th>Glenberg data</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$ Fluctuation parameter (sum of rates with which active and inactive elements become inactive and active, respectively)</td>
<td>0.087</td>
<td>0.082</td>
<td>0.013</td>
</tr>
<tr>
<td>$\beta$ Fluctuation parameter (ratio of the number of active elements to the total number of context elements)</td>
<td>0.288</td>
<td>0.150</td>
<td>0.260</td>
</tr>
<tr>
<td>$\gamma$ Scaling constant for context association</td>
<td>5.0\textsuperscript{a}</td>
<td>5.0\textsuperscript{a}</td>
<td>5.0\textsuperscript{a}</td>
</tr>
<tr>
<td>$\omega$ Probability that a new item enters the STS buffer</td>
<td>0.766</td>
<td>1.0\textsuperscript{b}</td>
<td>1.0\textsuperscript{b}</td>
</tr>
<tr>
<td>$\delta$ Amount of interitem information stored on a single study trial</td>
<td>0.688</td>
<td>0.246</td>
<td>0.732</td>
</tr>
<tr>
<td>$\alpha_2$ Constant representing the interfering effect of other memory traces in sampling</td>
<td>3.0</td>
<td>2.0</td>
<td>10.0</td>
</tr>
<tr>
<td>$\theta_2$ Scaling parameter in recovery equation for a test trial</td>
<td>0.5\textsuperscript{a}</td>
<td>0.300</td>
<td>0.215</td>
</tr>
<tr>
<td>$\lambda$ Rate of decay from STS</td>
<td>0.310</td>
<td>0.746</td>
<td>0.800</td>
</tr>
<tr>
<td>$L_{max}$ Maximum number of retrieval attempts</td>
<td>3\textsuperscript{a}</td>
<td>3\textsuperscript{a}</td>
<td>3\textsuperscript{a}</td>
</tr>
</tbody>
</table>

\textsuperscript{a} This parameter was not varied but kept fixed in the fitting of the model.

Graph of Raaijmakers (2003) model fits to results of Glenberg (1976) Experiment 1 results. Figure from Raaijmakers (2003) paper.

Fig. 3. Observed (left pane) and predicted (right pane) probabilities of recall as a function of spacing interval and retention interval (TR). Data from Glenberg (1976).
2.3.2 Fitting Young (1971) data

**FIGURE 2-3**

Graph of Pavlik and Anderson (2005) model fits to results of Young (1976) experiment. Figure from Raaijmakers (2003) paper.

![Graph of model fits to Young's data](image)

Fig. 2. Observed and predicted probabilities of recall as a function of spacing interval (number of intervening items). Data from Young (1971).

**FIGURE 2-3** details the results of the Raaijmakers (2003) model’s attempt to fit the data from the Young (1971) experiment. Parameters obtained for this fit are given in figure **FIGURE 2-1**, which resulted in a RMSE value of 0.686. Although this RMSE value appears really good, it is not surprising due to the limited number of data points being fitted by a relatively large number of parameters. Despite this concern, this fit to the Young (1971) data demonstrates that the Raaijmakers (2003) model can effectively account for non-monotonic performance.

2.3.3 Fitting Rumelhart (1967) data

**FIGURE 2-4** details the results of the Raaijmakers (2003) model’s attempt to fit the data from the Rumelhart (1967) experiment. Parameters obtained for this fit are given in figure **FIGURE 2-1**, which resulted in a RMSE value of 0.975 over the 8 lag sequence categories given. While this RMSE value is not as good as the one obtained for the Young (1971) data in **SECTION 2.3.2**, it should be kept in mind that the Rumelhart (1967) data consists of several more data points. Therefore, this fit further demonstrates that the Raaijmakers (2003) model can effectively account for non-monotonic performance, fitting to a larger amount of data than in the Young (1971) experiment.

2.3.4 Accounting for Ross & Landauer (1978)

It is argued in the Ross & Landauer (1978) publication, discussed in **SECTION 1.3.1.1**, that in encoding variability theories, the traces for the two study sessions are laid down independently of one another, and as a result, the probability of recall at evaluation should simply be the probability that *either* of the two traces are recalled.
Based on the introduction of a short-term memory buffer to SAM, detailed in SECTION 2.2.2, a critical assumption is made that, for a presentation of a given item, if the item is in long term memory AND is retrievable, the information from the current presentation is added to this original memory trace. As an example of applying this to the storing of contextual features, say at one instance the available contextual features are \( \{p, q, r, s\} \), and after a given amount of time these change to \( \{a, b, c, d\} \).
When one item is presented twice, massing these two presentations results in a neglect of the second presentation (since the first remains in short term memory, as detailed in SECTION 2.2.2) resulting in the formation of a single trace containing the \{p, q, r, s\} features. Likewise, when two items are each presented once, massing these two presentations results in the formation of 2 distinct traces, each containing the \{p, q, r, s\} features.

In contrast, when the two presentations are spaced for one item, the contextual features available at both presentations, \{p, q, r, s\} and \{a, b, c, d\}, are combined into one trace. However, in the case of two distinct items presented once, when these presentations are spaced out, they result in the formation of 2 distinct traces with different contextual elements. This situation is detailed in FIGURE 2-5.

<table>
<thead>
<tr>
<th>Contextual information available to massed vs. spaced items.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One item twice</strong></td>
</tr>
<tr>
<td>(p, q, r, s)</td>
</tr>
<tr>
<td><strong>Two items once</strong></td>
</tr>
<tr>
<td>(I_1)</td>
</tr>
<tr>
<td>(I_2)</td>
</tr>
</tbody>
</table>

2.4 General critique of Raaijmakers (2003) model

Despite the successes in fitting the data from the Glenberg (1976), Young (1971), and Rumelhart (1967) experiments, the Raaijmakers (2003) model still has several noteworthy challenges to overcome.

Although reasonably good at fitting experimental data, the Raaijmakers (2003) model requires 7 free parameters to do so (detailed in FIGURE 2-1). Although these free parameters are flexible enough to provide such fits to experimental data, it is also possibly they may be able to explain invalid data trends. In addition, while fits on data like that from Young (1971) are very good, due to the limited number of data points given, there is additional concern that the good fit may be due to parameter overfitting.

Regarding free parameter value distributions, it seems that some of them should be constant across over the 3 experiments, such as the context fluctuation parameters \(\alpha\) and \(s\). However, using the general-purpose optimization program (called MINUIT) to find optimal parameters, different values for these parameters are found for each experiment.

In both the fits to the Glenberg (1976) and Young (1971) data, a leveling off is observed which occurs after the optimal ISI is reached, which does not happen in the original SAM model. Out of the current problems the Raaijmakers (2003) model faces, it is this that is of most concern.
2.5 Meta analysis of Raaijmakers (2003)

Having analyzed the structure of the Raaijmakers (2003) model, and its performance in fitting experimental data, it becomes of interest to see how the model handles joint relationships between RI and ISI. Using the data from Glenberg (1976) experiment, and the parameter values chosen to fit this data (detailed in FIGURE 2-1), a log-log plot of the optimal ISIs for every RI in the data was conducted. The results of this plot are given in FIGURE 2-6 below.

FIGURE 2-6
Log-log plot of optimal ISI vs. RI, generated by model, for Glenberg (1976) data. Figure from Cepeda et al. (2006) paper.

This graph can be viewed in 3 separate sections. For very short RI values, very high optimal ISI values are preferred. This is likely due to the incorporation of the short-term memory buffer (introduced in SECTION 2.2.2). At these short RI values, items are more likely to be retained in short-term memory, making the first of two study trials mostly irrelevant. However, as RI values become high enough for items to leave short-term memory, a sharp decrease in optimal ISI occurs, reflecting a need for the material learned at the first study attempt. Once this threshold is reached, a sigmoidal curve is observed, increasing as a function of RI.

The results in FIGURE 2-6, generated by the Raaijmakers (2003) model, seem to contrast sharply with the empirical observations of this data in FIGURE 2-7 (provided
by Cepeda et al.) in which a steady increase in optimal ISI occurs as a function of RI. This suggests that major revisions to the Raaijmakers (2003) model may be necessary to properly explain joint relationships between RI and ISI.

**FIGURE 2-7**

*Plot of optimal ISI vs. RI, for empirical data from Cepeda et al. Figure from Cepeda et al. (2006) paper.*
3 Pavlik & Anderson model


3.1 Adaptive Character of Thought-Rational (ACT-R)

In ACT-R, for each rehearsal of a given memory image, it receives an increment of strength which decays as a function of time. Hence, the total strength of a given memory $m$ is denoted as the sum of all prior activations.

$$m(t_{1..n}) = \ln \left( \sum_{j=1}^{n} t_{j}^{-d} \right)$$  \hspace{1cm} \text{EQUATION 3-1}

In this equation, $n$ denotes the number of rehearsals for memory $m$, $t_{1..n}$ denote time (in seconds) since each rehearsal, and $d$ represents a fixed decay rate. Retrieval is dependent on the degree of activation, and will only occur if it is above a certain threshold, as defined below.

$$P_{\text{Recall}}(m) = \frac{1}{1 + e^{\frac{-m}{s}}}$$  \hspace{1cm} \text{EQUATION 3-2}

For the equation above, $\tau$ is defined as the threshold parameter and $s$ as the level of background noise in the memory activation. Through this equation, as the total strength $m$ of a memory image grows, the probability of its recall gradually approaches 1. Likewise, as threshold $\tau$ increases, the probability of recall gradually approaches 0. The $s$ parameter, denoting background noise, contributes a sensitivity factor to the equation. As the value of $s$ increases (e.g.: the system becomes noisier), the difference between threshold $\tau$ and total memory strength $m$ becomes of less consequence to the results of the equation.

3.2 Modifications to ACT-R

The objective of the Pavlik and Anderson (2005) model was to both predict both improvements in performance occurring with practice, and to predict decreases in performance occurring with delay. While the original ACT-R framework sufficed for these purposes, it failed to predict various results associated with the spacing effect, such as the interaction of spacing with retention intervals and with quantity of practice.

To allow for explanation of observable results associated with the spacing effect, Anderson and Schooler (1991) proposed a modification to the decay parameter $d$ in EQUATION 3-1. Rather than leaving decay as a fixed constant, it is instead declared as a time-based quantity:
\[ d_i(t_i, t_{i-1}) = \max[d, b(t_i - t_{i-1})^{-d}] \]  
\textit{EQUATION 3-3}

Where \( d \) represents a fixed minimum decay rate, and \( b(t_i - t_{i-1})^{-d} \) represents a decay value between time \( t_i \) and \( t_{i-1} \).

To allow for explanation of observable results associated with the spacing effect, an alternative approach was taken to the decay parameter \( d \) in \textit{EQUATION 3-1}. Namely, for the decay value on the \( i^{th} \) presentation given previous memory traces \( m(1-1) \), \( d_i(m_{i-1}) \), the following formula is used:

\[ d_i(m_{i-1}) = ce^{m_{i-1}} + \alpha \]  
\textit{EQUATION 3-4}

Where \( c \) denotes a scale factor on decay, and \( \alpha \) denotes an intercept factor. Using this alternative decay parameter, \textit{EQUATION 3-1} is modified accordingly:

\[ d_i(t_{1-n}) = \ln \left( \sum_{i=1}^{n} t_{i}^{-d} \right) \]  
\textit{EQUATION 3-5}

According to these equations, as the spacing increases between presentations, overall decay rate decreases -- resulting in a slower rate of forgetting for widely spaced learning.

3.3 Capabilities of the Pavlik & Anderson (2005) model

The Pavlik & Anderson (2005) publication detailed efforts to model experimental data from previous spacing effect experiments. Aside from modeling their own experimental data (see \textit{SECTION 1.2.3}), they attempted to model data from Experiment 1 in the Glenberg (1976) paper introduced in \textit{SECTION 1.2.1}, the experimental data from Young (1971) introduced in \textit{SECTION 1.2.2}, and the Rumelhart (1967) experiment introduced in \textit{SECTION 1.2.4}.

\textbf{FIGURE 3-1} provides a description of each of the parameters in the Pavlik and Anderson (2005) model, as well as values used for fitting data from the 3 experiments. To obtain these parameter fits, a general-purpose optimization program (known as MINUIT) was employed. This program worked by scanning over all the Pavlik and Anderson (2005) model parameters, given in \textbf{FIGURE 3-1}, and adjusting them to minimize the root mean squared error (\textit{RMSE}) statistic for a given set of experimental data.

\[ RMSE = \sqrt{\frac{\sum_{i=1}^{n} (obs_i - exp_i)^2}{n}} \]  
\textit{EQUATION 3-10}
In the $RMSE$ definition, $n$ is the number of data points and the values for $obs_i$ and $exp_i$ are the observed and predicted probabilities of successful recall for item $i$ out of $n$ items. The summation in the equation indicates a sum over all data points, and the division by $n$ gives the mean of the squared errors for the $n$ items.

### 3.3.1 Fitting Pavlik & Anderson (2005) experiment

**FIGURE 3-2** details the results of the Pavlik and Anderson (2005) model’s attempt to fit the data from the Pavlik and Anderson (2005) experiment. Parameters obtained for this fit are given in figure **FIGURE 3-1**, which resulted in an $RMSE$ value of 1.4714. This $RMSE$ value is not as good as those for the Glenberg (1976) Experiment 1, Young (1971), and Rumelhart (1967) experiments to be discussed shortly. However, despite this high $RMSE$ value, the Pavlik and Anderson (2005) model provides a very good qualitative fit to this experimental data.

### 3.3.2 Fitting Glenberg (1976) Experiment 1

**FIGURE 3-3** details the results of the Pavlik and Anderson (2005) model’s attempt to fit the data from Glenberg (1976), experiment 1. Parameters obtained for this fit are given in figure **FIGURE 3-1**, which resulted in an $RMSE$ value of 1.1529. This $RMSE$ value is not as good as those for the Young (1971) and Rumelhart (1967) experiments. In particular, for the Pavlik and Anderson (2005) model fits to 32 and 64 day RI curves (see **FIGURE 3-3**), noticeable flaws in the model fit can be observed.

### 3.3.3 Fitting Young (1971) experiment

**FIGURE 3-4** details the results of the Pavlik and Anderson (2005) model’s attempt to fit the data from the Young (1971) experiment. Parameters obtained for this fit are given in figure **FIGURE 3-1**, which resulted in an $RMSE$ value of 0.695. Although this $RMSE$ value appears really good, it is not surprising due to the limited number of data points being fitted by a relatively large number of parameters. Despite this concern, this fit to the Young (1971) data demonstrates that the Pavlik and Anderson (2005) model can effectively account for non-monotonic performance.

### 3.3.4 Fitting Rumelhart (1967) experiment

**FIGURE 3-5** details the results of the Pavlik and Anderson (2005) model’s attempt to fit the data from the Rumelhart (1967) experiment. Parameters obtained for this fit are given in figure **FIGURE 3-1**, which resulted in an $RMSE$ value of 0.9747 over the 8 lag sequence categories given. While this $RMSE$ value is not as good as the one obtained for the Young (1971) data in **SECTION 2.3.2**, it should be kept in mind that the Rumelhart (1967) data consists of several more data points. Therefore, this fit further demonstrates that the Raaijmakers (2003) model can effectively account for non-monotonic performance, fitting to a larger amount of data than in the Young (1971) experiment.
Table 1
Parameters and statistics for all data sets

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Decay intercept (a)</td>
<td>0.177</td>
<td>0.149</td>
<td>0.300</td>
<td>0.058</td>
</tr>
<tr>
<td>Decay scale (c)</td>
<td>0.217</td>
<td>0.495</td>
<td>0.419</td>
<td>0.283</td>
</tr>
<tr>
<td>Threshold (r)</td>
<td>-0.704</td>
<td>-0.704*</td>
<td>-0.704*</td>
<td>-0.704*</td>
</tr>
<tr>
<td>Noise (n)</td>
<td>0.255</td>
<td>0.255*</td>
<td>0.255*</td>
<td>0.255*</td>
</tr>
<tr>
<td>Interference scalar (b)</td>
<td>0.025</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Encoding scalar (b)</td>
<td>1*</td>
<td>1*</td>
<td>1*</td>
<td>0.352</td>
</tr>
<tr>
<td>Reduced encoding (b_f)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>0.274</td>
</tr>
</tbody>
</table>

Fit Statistics
- $r^2$: 0.944 0.927 0.461 0.944
- RMSD adjusted: 0.046 0.021 0.026 0.026
- $\chi^2$: 328 41.5 8.70 31.9
- $\chi^2 df$: 157 38 16 20

Note. RMSD = root mean square deviation.
*Fixed parameters.
FIGURE 3-3
Figure from Pavlik and Anderson (2005) paper.

FIGURE 3-4
Graph of Pavlik and Anderson (2005) model fits to results of Young (1976) experiment. Figure from Pavlik and Anderson (2005) paper.
FIGURE 3-5
Graph of Pavlik and Anderson (2005) model fits to Rumelhart (1967) results. Figure from Pavlik and Anderson (2005) paper.

(a) 10-10-10-10-10 Spacing
(b) 10-10-1-1-10 Spacing
(c) 6-6-6-6-10 Spacing
(d) 6-6-1-1-10 Spacing
(e) 3-3-3-3-10 Spacing
(f) 1-1-10-10-10 Spacing
(g) 1-1-6-6-10 Spacing
(h) 1-1-1-1-10 Spacing
2.4 General critique of Pavlik & Anderson (2005) model

In general, the Pavlik and Anderson (2005) model was able to produce relatively good qualitative fits to the data from the Pavlik and Anderson (2005), Glenberg (1976) Experiment 1, Young (1971), and Rumelhart (1967) experiments. Nonetheless, there were certain situations in which the fits were off. In particular, for the Pavlik and Anderson (2005) model fits to the Glenberg (1976) Experiment 1 32 and 64 day RI curves (see FIGURE 3-3), noticeable flaws in the model fit can be observed.

The Pavlik and Anderson (2005) model used 2 free parameters to fit data (the decay intercept $\alpha$ and decay scale $c$, detailed in FIGURE 3-1). This leads to a considerably more concise model than the Raaijmakers (2003) model (which relied upon 7 free parameters). While a model this concise provides for more concise fits to experimental data, it may not be descriptive enough to explain certain data patterns, as in Glenberg (1976) Experiment 1 (see FIGURE 3-3).

2.5 Meta analysis of Pavlik & Anderson (2005) model

Having analyzed the structure of the Pavlik and Anderson (2005) model, and its performance in fitting experimental data, it becomes of interest to see how the model handles joint relationships between retention interval and inter-study interval.

The objective of this meta-analysis was to fit the Pavlik and Anderson (2005) model to the empirical observations of optimal inter-study interval to retention interval pairs provided by Cepeda et al. (2006). Using the optimized parameters derived for four sets of experimental data (see FIGURE 3-1), and the parameter values chosen to fit this data (detailed in FIGURE 3-1), a log-log plot of the optimal ISIs for every RI in the data was conducted.

FIGURE 3-7 gives a graph of the log-log plot of optimal ISI to RI by Pavlik and Anderson (2005) model for Pavlik and Anderson (2005) data, which produces a slope of .8870. FIGURE 3-8 provides a graph of the log-log plot of optimal ISI to RI by Pavlik and Anderson (2005) model for Rumelhart (1967) data, which produces a slope of .8956. FIGURE 3-9 gives a graph of the log-log plot of optimal ISI to RI by Pavlik and Anderson (2005) model for Young (1967) data, which produces a slope of .8138. FIGURE 3-10 Graph of log-log plot of optimal ISI to RI by Pavlik and Anderson (2005) model for Glenberg (1976) Experiment 1 data, which produces a slope of .9164.

With the parameter sets for each experiment, a log-log slope of less than 1 was observed. Going back to SECTION 1.2.5.3, it is important to note that this indicates the absence of a fixed ratio between optimal inter-study interval and retention interval. These findings reinforce doubt on whether a special ratio exists between optimal inter-study interval and retention interval.
FIGURE 3-6
Plot of optimal ISI vs. RI, for empirical data from Cepeda et al. Figure from Cepeda et al. (2006) paper.

FIGURE 3-7
FIGURE 3-8

FIGURE 3-9
Graph of log-log plot of Optimal ISI to RI by Pavlik and Anderson (2005) model for Young (1967) data.
FIGURE 3-10

SLOPE = 0.3164    RMSE = 0.8216
4 The Extended TAC-V model

In this paper, we present a new model based on the Temporal Associative Context Variability model, first described in Cepeda, Mozer, Coburn, Rohrer, Wixted, and Pashler (unpublished).

4.1 The original TAC-V model

The original variant of the Temporal Associative Context Variability (TAC-V) model was first described in Cepeda, Mozer, Coburn, Rohrer, Wixted, and Pashler (unpublished). Like the Search of Associative Memory (SAM) and Raaijmakers (2003) models, the TAC-V model relies heavily upon encoding variability theory, described in detail in SECTION 1.3.1.

As a brief refresher, encoding variability theory states that each time an item is stored in memory, it is stored with information from the learning context. This information consists of discrete features such as room temperature, brightness of the surrounding walls, and so on. As time passes, the discrete elements of this context change, resulting in a temporal fluctuation process. For retrieval of the originally learned item to occur, the context during retrieval must be sufficiently similar to context at learning. In order for TAC-V to represent this context, an $n$-dimensional vector $R^n$ is relied upon, where $R$ denotes a particular instance in time, and where each component in $R \{1…n\}$ represents a discrete value for a contextual feature in the environment.

In order to model contextual variations, TAC-V relies upon 6 variables:
- $c_{S_1}$: The context representation at the start of learning opportunity $S_1$
- $c_{S_2}$: The context representation at the start of learning opportunity $S_2$
- $c_T$: The context representation at the start of the evaluation session $T$
- $ISI$: The time lag between the end of $S_1$ and the start of $S_2$
- $RI$: The time lag between the end of $S_2$ and the start of $T$
- $R$: The probability of recalling at the evaluation $T$, based on possible recall of $S_1$ or $S_2$

According to encoding variability theory, the probability of recall $R$ is dependent on the similarity of the context at $T$ to that of either study session $S_1$ or $S_2$. As such, recall $R$ is dependent on context representations $c_T$, $c_{S_1}$, and $c_{S_2}$. FIGURE 4-1 illustrates a random walk for one contextual element $n$ between the contexts $c_{S_1}$, $c_{S_2}$, and $c_T$ as a function of time. One observation is that, since the context value is equally likely to increase or decrease, its variance (or the range of places it can be in the future) grows linearly with time. Another observation is that, for any given instant, the value of context is dependent on its previous locations, where the value of $c_T$ depends on $c_{S_2}$, which in turn depends on $c_{S_1}$. In addition, because variance is time-dependent, the value of $c_{S_2}$ depends on $ISI$. 

42
and \( c_r \) depends on \( RI \). The dependencies between these variables are expressible in terms of a Bayesian framework, as illustrated in **FIGURE 4-2**.

**FIGURE 4-1**

*Example of one-dimensional contextual random walk for TAC-V*

**FIGURE 4-2**

*Bayesian framework for the original TAC-V model*
Because encoding variability theory is only concerned with variations between these contexts, the value $c_{S_i}$ becomes irrelevant (for modeling purposes, TAC-V sets it to 0 in all cases). To determine the value of $c_{S_2}$, EQUATION 4-1 is used.

$$P(c_{S_2} \mid c_{S_1}, ISI) = N(c_{S_2} ; c_{S_1}, \alpha ISI)$$ \hspace{1cm} \text{EQUATION 4-1}$$

In this formula, the value of $c_{S_2}$ is derived from a gaussian density function with a mean value of $c_{S_1}$ and variance of $\alpha ISI$, where $\alpha$ is a free parameter for scaling variance. The gaussian density function used takes the form given in EQUATION 4-2, in which independence of each individual contextual feature $n$ in the vector $R^n$ is taken into account.

$$N(x; \mu, \sigma^2) = \frac{1}{(2\pi \sigma^2)^{\frac{n}{2}}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ \hspace{1cm} \text{EQUATION 4-2}$$

Likewise, to determine the value of the context at $T$, a formula like EQUATION 4-1 is used, in which $c_{S_2}$ is the mean value and $\alpha RI$ is the variance. This is presented in EQUATION 4-3 below.

$$P(c_{T} \mid c_{S_2}, RI) = N(c_{E} ; c_{S_2}, \alpha RI)$$ \hspace{1cm} \text{EQUATION 4-3}$$

To determine probability of recall $R$ at evaluation $T$, based on either $c_{S_1}$, with $S_i$ equaling either $S_1$ or $S_2$, EQUATION 4-4 is used.

$$P(R \mid c_{T}, c_{S_i}) = e^{-\rho |c_{T} - c_{S_i}|}$$ \hspace{1cm} \text{EQUATION 4-4}$$

This equation assumes an exponential decrease in probability of recall as a function of the squared distance between context at $c_{S_i}$ and $c_{T}$. The free parameter $\rho$ represents a scaling factor on recall. Since recall depends on similarity of $c_{T}$ to either $c_{S_1}$ or $c_{S_2}$, recall succeeds if either of these are recalled. Therefore, the probability of recall $R$ at evaluation $T$ is expressed by EQUATION 4-5 below.

$$P(R \mid c_{T}, c_{S_1}, c_{S_2}) = P(\text{recalling } S_1 \text{ OR recalling } S_2) = 1 - P(\text{NOT recalling } S_1 \text{ OR recalling } S_1) = 1 - [P(\text{NOT recalling } S_1)][P(\text{NOT recalling } S_2)] = 1 - [1 - P(\text{recalling } S_1)][1 - P(\text{recalling } S_2)] = 1 - [1 - P(R \mid c_{T}, c_{S_1})][1 - P(R \mid c_{T}, c_{S_2})]$$ \hspace{1cm} \text{EQUATION 4-5}$$
In the form presented here, the TAC-V model only has 3 free parameters: the number of contextual elements \( n \), the linear scaling factor on variance \( \alpha \), and the scaling factor on recall \( \rho \). However, on close inspection, it appears that the parameters \( n \) and \( \alpha \) are redundant. In other words, increasing the value of \( n \) has the same effect as decreasing the value \( \alpha \), and vice-versa. This is observable in the exponential component of the Gaussian density function from Equation 4-2.

\[ \frac{\left| c_i - c_j \right|^2}{\alpha \Delta t} \]  

\text{Equation 4-6}

For the context representations \( c_i \) and \( c_j \), their distance has a linear relationship with the number of context elements \( n \). For instance, this distance between \( c_i \) and \( c_j \) for a single contextual element can be the equivalent to the sum of smaller distances for multiple contextual elements. This linear relationship with the number of context elements \( n \) results in a cancellation with the \( \alpha \) parameter in the denominator. Therefore, for modeling purposes, TAC-V sets \( n \) to be constant at 1.

In addition to the redundancy between the linear scaling factor on variance \( \alpha \) and the number of contextual elements \( n \), it appears that \( \alpha \) and the scaling factor on recall \( \rho \) are redundant as well, because these variables each have a linear relationship to the exponent of Equation 4-4. To verify this redundancy, we used this version of TAC-V to fit the experimental data of Cepeda et al. (2006), using combinations of \( \alpha \) values ranging from \( 2^{-6} \) to \( 2^{16} \) and \( \rho \) values ranging from \( 2^{-18} \) to \( 2^{4} \) (the specific values used are given in Figure 4-3). For each value of \( \rho \), a search over all possible values \( \alpha \) was done to find the parameter value that would result in the best model fit to the experimental data. A graph of these results is given in Figure 4-4, which shows a clear linear relationship between \( \alpha \) and \( \rho \).

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 2^{-16} )</td>
<td>( 2^{-18} ), ( 2^{-16} ), ( 2^{-14} ), ( 2^{-12} ), ( 2^{-10} ), ( 2^{-8} ), ( 2^{-6} ), ( 2^{-4} ), ( 2^{-2} ), ( 2^{0} ), ( 2^{2} ), ( 2^{4} )</td>
</tr>
</tbody>
</table>

**Figure 4-3**

Range of \( \alpha \) and \( \rho \) values searched to check their redundancy in original TAC-V model.
The objective of the framework presented is to provide a means of estimating the values given by EQUATION 4-7 below,

\[ P(R \mid ISI, RI) = \int_{c_{S_1}}^{c_{S_2}} \int_{c_y} N(c_{S_2}, 0, aISI) N(c_T; c_{S_2}, aRI) P(R \mid c_T, 0, c_{S_2}) \]

which integrates recall probability over all possible representations of context at S2 and E, weighted by the probability of obtaining those representations.

### 4.1.1 Meta-analysis of the TAC-V model

While EQUATION 4-7 cannot be analytically solved, it is possible to obtain an approximation to its solution by finding optimal inter-study intervals – those that maximize recall probability – for a given range of retention intervals. This was attempted in Cepeda, Mozer, Coburn, Rohrer, Wixted, and Pashler (unpublished), where the objective was to use TAC-V fit human data derived from meta analysis conducted by Cepeda et al., shown in FIGURE 1. By arbitrarily setting \( \rho = 1 \), a linear log-log plot was generated with a slope of 0.673, as shown in FIGURE 2.
This successfully reproduced two qualitative patterns present in the data from Figure 1. As RI increases, so does optimal ISI. As RI increases, the optimal ISI to RI ratio decreases (indicated by the slope of the TAC-V fit being less than 1).

**FIGURE 4-5**

Log-log plot of optimal ISI value by RI, for all studies in the Cepeda, Mozer, Coburn, Rohrer, Wixted, and Pashler (unpublished) meta-analysis for which the optimal ISI was flanked by shorter and longer ISIs. The dashed line shows the best fit power regression line for the observed data.
4.2 Efforts taken to improve the TAC-V model

Looking at the structure of the original TAC-V model, only three free parameters are given: the number of contextual elements \( n \), the linear scaling factor on variance \( \alpha \), and the scaling factor on recall \( \rho \). The parameters \( n \) and \( \alpha \) become redundant with each other based on the relationship in EQUATION 4-6. In this equation, the context representations \( c_i \) and \( c_j \) have a linear relationship with the number of context elements \( n \), resulting in a cancellation with the \( \alpha \) parameter in the denominator. In addition, the parameters \( \alpha \) with \( \rho \) based on the relationship given in FIGURE 4-4.

Inherently, every experimental study varies in detailed methodology, for things such as the type of stimuli used (e.g., verbal versus visual, concrete vs. abstract), the manner of testing, and so on. Due to this variability, it is not possible to explain these studies with only these three redundant parameters in the model. Although this doesn't suggest a weakness in the model, it does signify that TAC-V lacks the expressiveness required to account for experimental data.
4.2.1 Variant 1 of TAC-V: Incorporating fixed encoding probability

According to TAC-V, memory encodings for a given item are assumed to take place at $S_1$ and $S_2$. However, it is possible that the information fails to encode in memory at either $S_1$ or $S_2$. Because of possible failure to encode information, it becomes practical to incorporate parameters to represent encoding probability. For $c_{S_1}$ and $c_{S_2}$, the probabilities of encoding can be denoted by the fixed variables $E_1$ and $E_2$, respectively. To represent the probability of successful encodings (that is $P(E_1 = \text{TRUE})$ or $P(E_2 = \text{TRUE})$), another free parameter $\varepsilon_1$ is incorporated, ranging from 0 for difficult items to 1 for easy items. Incorporation of these encoding parameters results in the following modifications to \textbf{EQUATION 4-5}, given below in \textbf{EQUATION 4-8}.

\[ P(R \mid E_1, E_2, c_T, c_{S_1}, c_{S_2}) \]
\[ = 1 - [(1 - P(R \mid c_T, c_{S_1}))^{E_1}][(1 - P(R \mid c_T, c_{S_2}))^{E_2}] \]

\textbf{EQUATION 4-8}

Where $P(E_1=\text{TRUE}) = P(E_2=\text{TRUE}) = \varepsilon$ for a successful encoding, and $P(E1=\text{FALSE}) = P(E2=\text{FALSE}) = (1 - \varepsilon)$ for a failed encoding.

These modifications to the original TAC-V framework are presented in the Bayesian network given in \textbf{FIGURE 4-7} below.

\textbf{FIGURE 4-7}

Bayesian network of TAC-V Variant 1 (incorporating fixed encoding probability)
4.2.2 Variant 2 of TAC-V: Incorporating variable difficulty of item encoding

In learning, certain kinds of information will be inherently harder to commit to memory. For instance, most individuals may have an easier time learning French-English foreign language pairs than learning names of complex chemical compounds. It is therefore appropriate to assume that certain items will have lower probabilities of encoding than others, as opposed to one fixed encoding difficulty $\varepsilon$, as presented in variant 1 of TAC-V.

One key motivation for incorporating varying difficulty of items is to resolve the argument posed by the Ross and Landauer (1978) paper, described in detail in SECTION 1.3.1.1. They argued that in encoding variability theories, the traces for the two study sessions A and B are laid down independently of one another, and as a result, the probability of recall at evaluation should simply be the probability that either the A or B trace is recalled. However, this argument basically assumes that items A and B have equal difficulties, which result in equal encoding and recall probabilities. If A and B have different encoding difficulties, different probabilities will result for recalling one A studied twice or either A or B each studied once.

Deriving the values of encoding variables $E_1$ and $E_2$ from some distribution of item difficulties results in a modified Bayesian network given in FIGURE 4-8 below.

![Bayesian network of TAC-V Variant 2 (incorporating fixed encoding probability)](image-url)
4.2.3 Variant 3 of TAC-V: Incorporating conditioning on recall at $S_2$

Another issue to consider is that, depending on whether or not recall occurs at learning opportunity $S_2$, it can impact whether encoding occurs for $S_2$. An example of this is in the experimental data from Cepeda et al. (2006), first mentioned in SECTION 1.2.4.2, shown in FIGURE 4-9 for the fact learning section of the experiment, and FIGURE 4-10 for the picture learning. For both of these graphs, the black dashed line indicates the probability of recalling information from learning session $S_1$ at the start of session $S_2$.

Likewise, the blue line represents probability of recalling information from $S_1$ or $S_2$ at the start of the evaluation session $T$. In addition to these, the green line represents the probability of recall at evaluation $T$ given successful recall of information at learning opportunity $S_2$, whereas the red line represents failure to recall at $S_2$. In both the fact and picture learning experiments, it is clear that the green curve is considerably higher than the red, such that $P(R_E \mid R_{S2} = \text{TRUE}) > P(R_E \mid R_{S2} = \text{FALSE})$.

Relating back to the concept of item difficulty, when an easy item is presented, it is likely to be recalled at $S_2$, and so recall at $S_2$ is predictive of the fact it’s an easy item and will more likely be recalled at test.

Incorporating a dependency for $E_2$ on recall of information from $S_1$ at results in a modified Bayesian network given in FIGURE 4-11 below. Based on the arguments made through the experimental data in Cepeda et al. (2006), it was decided that this variant of TAC-V would be focused on over variants 1 and 2 presented earlier.

**FIGURE 4-9**
Experimental data obtained from the fact learning section (Experiment 2a) of the Cepeda et al. (2006) publication (described in SECTION 1.2.4.2)
Experimental data obtained from the picture learning section (Experiment 2b) of the Cepeda et al. (2006) publication (described in SECTION 1.2.4.2)

FIGURE 4-10

Bayesian network of TAC-V Variant 2 (incorporating fixed encoding probability)
4.2.4 Addressing how to incorporate difficulty

In the previous section, variants of the TAC-V model were proposed which incorporate some form of varying difficulty for items. With these structures defined, the issue of mathematically obtaining varying item difficulties can now be explained. One method we explored for doing this was introducing a distribution over item difficulty via encoding probability. An alternative approach we took for this involved introducing a distribution over item difficulty via modulation of contextual distance parameters.

4.2.4.1 Item difficulty via encoding probability

For this approach, values of difficulty were pulled from a Beta distribution, of the form

\[
\text{Beta}(x;\varepsilon_1,\varepsilon_2) = \frac{1}{B(\varepsilon_1,\varepsilon_2)} x^{\varepsilon_1-1} (1-x)^{\varepsilon_2-1}
\]

\[\text{EQUATION 4-9}\]

\[
\text{Mean} = \frac{\varepsilon_1}{\varepsilon_1 + \varepsilon_2} \quad \text{Variance} = \frac{\varepsilon_1\varepsilon_2}{(\varepsilon_1 + \varepsilon_2)^2 (\varepsilon_1 + \varepsilon_2 + 1)}
\]

Where the variables \(\varepsilon_1\) and \(\varepsilon_2\) denote shaping parameters, and the function \(B(\varepsilon_1,\varepsilon_2)\) is a normalization constant on the Beta distribution itself.

4.2.4.1.1 Fitting Cepeda et al. (2006) data

In order to test the effectiveness of adding the encoding probability approach to the TAC-V model, an attempt was made to fit the experimental data from the Cepeda et al. (2006) paper for both the fact learning and picture learning sections. To do this, a series of simulations were conducted using different values of mean (ranging from .25 to 1) and variance (ranging from .5 to 2.5) for the beta distribution (derived from the shaping parameters \(\varepsilon_1\) and \(\varepsilon_2\), through EQUATION 4-9). The purpose of conducting these simulations was to find parameter values which would allow the model to best fit to the experimental data for the probability of recalling information from \(S_1\) or \(S_2\) at the start of the evaluation session \(T\) (denoted by the blue line in FIGURE 4-9 and FIGURE 4-10).

Results of this attempt are given in FIGURE 4-11 for the fact learning section of the experiment, and FIGURE 4-12 for the picture learning section. For fact learning, the optimal beta distribution mean was .5, and the optimal variance was .5, resulting in a root mean squared fitting error of 0.02823. Likewise, for picture learning, the optimal beta distribution mean was .5, and the optimal variance was 1.5, resulting in a root mean squared fitting error of 0.04133.
Fit of item difficulty via encoding probability variant of TAC-V to the fact learning section (Experiment 2a) of the Cepeda et al. (2006) publication (described in SECTION 1.2.4.2). In the bottom plot, the magenta curve denotes the experimental data values, and the blue curve denotes the model fit.

![Graph showing fit of item difficulty via encoding probability variant of TAC-V to the fact learning section.](image-url)
Fit of item difficulty via encoding probability variant of TAC-V to the picture learning section (Experiment 2b) of the Cepeda et al. (2006) publication (described in SECTION 1.2.4.2). In the bottom plot, the magenta curve denotes the experimental data values, and the blue curve denotes the model fit.
4.2.4.1.2 Problems with this approach

In the attempt to fit experimental data from Cepeda et al. (2006), two problems became apparent. In both FIGURE 4-9 and FIGURE 4-10, the probability of recall at evaluation T given successful recall of information at learning opportunity S2, represented by the green curve, should have a sharper rise than it does. In addition, while the model fit to the data seems to capture its general trend, such as having a global peak in the right place, it does not capture much of the data trend beyond this.

4.2.4.2 Item difficulty via modulation of contextual distance parameters

As an alternative to denoting item difficulty through encoding probability, using modulations on contextual distance appears promising. One way to think about difficulty is that easier items are more robust to the context at test wandering away from the context at study, whereas more difficult items are more sensitive to the context.

To incorporate modulation of contextual distance parameters, EQUATION 4-4 in the original TAC-V model is modified to include 2 new parameters: $\gamma$, which serves as a scaling factor occurring outside of the exponent, and $\eta$, which serves as a scale on the degree of magnitude of the exponent. These modifications are presented in EQUATION 4-10 below.

$$P(R \mid c_E, c_S) = \gamma e^{-\rho |c_E - c_S|}$$

EQUATION 4-10

In addition to these parameters, an additional parameter is needed, occurring outside the term $P(R \mid c_E, c_S)$, to offset the curve values by a certain amount. Although this does nothing to alter the curve trends, it appears to be necessary for getting the curves to the same height as the experimental data. This offset parameter, $\phi$, is factored in as such:

$$P(R \mid c_E, c_S) = \phi + P(R \mid c_E, c_S)$$

EQUATION 4-11

4.2.4.2.1 Fitting Cepeda et al. (2006) data

In order to test the effectiveness of adding the modulation of contextual distance parameters to the TAC-V model, an attempt was made to fit the experimental data from the Cepeda et al. (2006) paper for both the fact learning and picture learning sections. To do this, a series of simulations were conducted using different values of $\gamma$ (ranging from 1 to 25), $\rho$ (drawn from a standard gaussian distribution, with mean ranging from -10 to 10, and variance ranging from 0 to 10), and $\eta$ (ranging from 0 to 3). The purpose of this was to find values for these which would allow the model to best fit to the experimental data for the probability of recalling information from $S_1$ or $S_2$ at the start of the evaluation session T (denoted by the blue line in FIGURE 4-9 and FIGURE 4-10).
Results of this attempt are given in FIGURE 4-14 for the fact learning section of the experiment, and FIGURE 4-15 for the picture learning section. For fact learning, the optimal beta Gaussian distribution mean for deriving $\rho$ was $2^{-4.99472}$ with an optimal variance was 0.39313, the optimal value of $\eta$ was 0.738209, the optimal $\gamma$ was 25, and the optimal offset $\phi$ was -2.47457, resulting in a root mean squared fitting error of 0.03819. Likewise, for picture learning, the optimal beta Gaussian distribution mean for deriving $\rho$ was $2^{-4.96436}$ with an optimal variance was 0.76925, the optimal value of $\eta$ was 0.63506, the optimal $\gamma$ was 25, and the optimal offset $\phi$ was -3.294240, resulting in a root mean squared fitting error of 0.0701.

**FIGURE 4-14**

*Fit of item difficulty via the modulation of contextual distance parameters variant of TAC-V to the fact learning section (Experiment 2a) of the Cepeda et al. (2006) publication (described in SECTION 1.2.4.2). The blue curve denotes the experimental data values, and the red curve denotes the model fit.*
4.2.4.2.2 Fitting Glenberg data

In order to test the effectiveness of adding the modulation of contextual distance parameters to the TAC-V model, an attempt was made to fit the experimental data from the Glenberg (1979) paper for both the fact learning and picture learning sections. In contrast to the data from the Cepeda et al. (2006) publication, this data consists of 4 different retention intervals (hence, 4 curves). Therefore, it becomes of interest to see whether parameter values can be found which provide reasonable fits over all values of retention interval.

To do this, a series of simulations were conducted using different values of $\gamma$ (ranging from 1 to 25), $\rho$ (drawn from a standard gaussian distribution, with mean ranging from -10 to 10, and variance ranging from 0 to 10), and $\eta$ (ranging from 0 to 3). The purpose of this was to find values for these which would allow the model to best fit to the experimental data for the probability of recalling information from $S_1$ or S2 at the start of the evaluation session T (denoted by the blue curves in FIGURE 4-16).

Results of this attempt are given in FIGURE 4-16, shown by the red curves. For fact learning, the optimal beta Gaussian distribution mean for deriving $\rho$ was $2^{0.134141}$ with an optimal variance was 5.183712, the optimal value of $\eta$ was 1.499400, the optimal $\gamma$ was 20.798343, and the optimal offset value was -4.894598, resulting in a root mean squared fitting error of 0.08851.
FIGURE 4-16

Fit of item difficulty via the modulation of contextual distance parameters variant of TAC-V to the data from Experiment 1 of the Glenberg (1979) publication (described in SECTION 1.2.1). The blue curves denote the experimental data values, and the red curves denote the model fit.
Concluding Remarks

In recent decades, a handful of efforts have been made to explain the spacing effect through cognitive modeling (Raaijmakers (2003), Pavlik & Anderson (2005), Young (1971), Cepeda, Mozer, Coburn, Rohrer, Wixted, and Pashler (unpublished)). While these efforts appear promising, they each contain a unique set of flaws, preventing a full and concise explanation of the spacing effect.

The objective of our work for this project was twofold. First, we intended to analyze previous efforts in modeling the spacing effect. Following this, we intended to extend the Temporal Associative Context Variability (TAC-V) framework presented in Cepeda, Mozer, Coburn, Rohrer, Wixted, and Pashler (unpublished) to model previously collected data from experimental literature (Pavlik and Anderson (2005), Glenberg (1976), Cepeda et al. (2006)).

The original TAC-V model provided a very concise framework for fitting data from the experimental literature, relying on 1-2 free parameter(s) (SEE SECTION 4-1). Because of this, we believed that TAC-V lacked the expressiveness to effectively account for patterns in data. To remedy this, we explored two ways of extending the TAC-V model, each incorporating the concept of variable difficulty of items to be encoded (SEE SECTION 4-1).

For the TAC-V model itself, there still remain a number of possible things to consider for future work. One possible reason encoding at S2 may fail lies in consolidation theory, where if the ISI is too short, S2 interferes with the encoding of $S_1$. Another possible reason this may happen lies in inattention theory, detailed in SECTION 1.3.2, where failure to attend to S2 occurs if ISI is too short.

Currently, the underlying mechanisms of the spacing effect remain unresolved. Several theories have been proposed to account for them, such as encoding variability, introduced in SECTION 1.3.1, and inattention theory, introduced in SECTION 1.3.2. However, neither of these mechanisms have gained universal acknowledgement. Encoding variability, relied upon by the majority of spacing effect models, still faces unresolved challenges posed by the Ross and Landauer (1978) publication.

In addition to theoretical issues, research into the spacing effect continues to face empirically oriented issues. While a large number of experiments have been conducted, the majority required subjects to remember information for less than 2 weeks before final evaluation of knowledge. In the majority of real-world situations, such as with classroom-based education, retention of knowledge is expected to be much longer than 2 weeks. As such, little practical application of the spacing effect (to things such as classroom lesson planning) has taken place so far. The empirical data provided by Cepeda et al. (2006) (SEE SECTION 1.2.5.2) provides some insight into the spacing effect over long term retention intervals (6 months in length). However, for further empirical analysis of long term spacing effects to occur, more experiments relying upon long retention intervals must be conducted.
References


