INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9417498

A sparse distributed memory model of overregularization

Nicols, Annemarie, Ph.D.

City University of New York, 1994

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A SPARSE DISTRIBUTED MEMORY MODEL OF OVERREGULARIZATION

by

ANNEMARIE NICOLS

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1994

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

1-24-94
Date

Martin Chodorow
Chair of Examining Committee

1-24-94
Date

Executive Officer

Martin Chodorow

Virginia Valian

Donald Scarborough

David Palermo

Eric Heinemann
Supervisory Committee

The City University of New York

Abstract

A SPARSE DISTRIBUTED MEMORY MODEL OF OVERREGULARIZATION

by

Annemarie Nicols

Advisor: Professor Martin Chodorow

As children acquire language they often make mistakes. One of the most notable of these errors is the overregularization (OR) of irregular past tense verbs since numerous studies have indicated that past tense acquisition is U-shaped (Cazden, 1968; Ervin, 1964; Kuzjac, 1977; Slobin, 1971). Children start out using the correct form of the verb, (e.g. broke, came) then go through a stage of overregularization, producing incorrect forms (e.g. breaked, comed) and finally, they gradually return to the use of the correct form. It may be argued that U-shaped learning and OR are general cognitive phenomena since a number of studies have shown that they occur when adults try to learn rule based systems with exceptions (Palermo & Eberhart, 1968; Palermo & Howe, 1970) and in other types of problem-solving tasks (Karmiloff-Smith & Inhelder, 1974/75).

Traditionally, OR and its U-shaped developmental sequence have been explained with a two process theory that involves both rote memory and rule use. In contrast, recent interest in connectionist theories of human memory have offered one process theories of OR which do not require the use of explicitly defined rules (MacWhinney & Leinbach, 1991; Plunkett & Marchman, 1991, 1993; Rumelhart & McClelland, 1986). The success of these connectionist models at simulating U-shaped learning and OR have led to recent modifications of the traditional theory (Pinker,1991; Marcus et al., 1992).

In 1988 Kanerva introduced sparse distributed memory (SDM), a mathematical memory model, inspired by the biology of the human nervous system. This dissertation offers evidence that, although different from the connectionist models in a number of ways, SDM can account for OR and its U-shaped developmental sequence without built-in rules. This evidence includes: 1) a SDM simulation of Rumelhart and McClelland's (1986) "rule of 78", which is a simple demonstration of OR and U-shaped development, 2) a SDM simulation of Palermo and Howe's (1970) non-linguistic OR data, 3) a paired associates learning experiment with human subjects which was designed to investigate the effect of "friends" and "enemies" on OR and its learning curve, and lastly 4) a SDM simulation of this experiment.

ACKNOWLEDGEMENTS

To my advisor, Martin Chodorow:

Thank you for introducing me to SDM and for all your suggestions, support and guidance over these many years. I look forward to working together in the future to see what SDM can really do.

To my committee, Virginia Valian, Donald Scarborough, Eric Heinemann and David Palermo:

Thank you for all your help and suggestions. Your contributions helped to make this a work I can be proud of.

To my parents, Angela and Otto:

Thank you, first and foremost, for being the most wonderful parents the five of us could ever hope for. Your love, support, encouragement and belief in me and my abilities helped me through this (and so many other things) more than you'll ever know. I love you both very much.

To my sister and brothers, Beth, John, Bill and Rich:

Thanks for all the fun and laughter at those times when I needed it most. I am a very lucky sibling ! I love you all.

And to Peter:

Thank you for making me happy enough to finally finish this.

Thank you for your unconditional support and belief in me. It was often the only thing that kept me going through the tough times.

Thank you for being the person you are. Your intelligence and talent allow you the opportunity to do exceptional things, but your determination and hard work get them done. By your example, I have learned that I can do anything I set out to do.

And most importantly, thank you for filling my life with love, fun and happiness. **I love you...always and forever.**

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. Introduction

As children acquire language they often make mistakes. One of the most notable of these errors is the overregularization of irregular past tense verbs. A child may say "breaked" instead of "broke" or "comed" instead of "came" to indicate the past tense of "to break" and "to come" respectively. It appears as if the child is over-applying the regular "stem + ed" rule for indicating the past tense, to irregular verbs, thus the term overregularization (OR). Not only do children invariably overregularize, they have been reported to do so in a specific developmental sequence (Cazden, 1968). At first, children inflect irregular verbs correctly, if they do so at all. This is followed by a period during which the child makes some overregularization errors. Finally, the overregularizations decrease to a point where they are extremely rare. This decrease in performance, followed by a subsequent increase, results in a U-shaped development curve when the percentage of correctly marked irregular verbs is plotted against age. Overregularization and its interesting developmental sequence have received considerable attention because they are often regarded as general cognitive phenomena, not exclusive to language acquisition. This attention has spurred the development of numerous theories to explain the phenomena and these theories have taken one of two approaches. The major theories will be discussed in detail later, however it seems prudent to point out the major differences between the approaches at this point. Both the traditional theory and the Blocking and Memory Retrieval Failure theory proposed by Marcus, Pinker, Ullman, Hollander, Rosen and Xu (1992) account for overregularization using two distinctly separate processes, rote memory and the use of a rule system . In contrast, the theories which result from computer simulations of the phenomenon, specifically the Rumelhart and McClelland (1986) parallel distributed processing (PDP) model and Kanerva's (1988) sparse distributed memory (SDM) model account for the phenomenon in systems which require only a single process, without explicitly defined rules. The way each of these divergent theories handles both the linguistic and nonlinguistic OR data will be a major theme of this work. This introduction will begin with a review of the language acquisition literature in order to document the existence of the phenomena in this domain. The "two process" theories, since they follow a more traditional cognitive approach to the phenomenon, will then be examined. Next, the nonlinguistic evidence for overregularization will be presented and finally, the "one process" theories, which have arisen as a result of computerized simulations of OR, will be introduced.

A. Overregularization in Language Acquisition

In an early study, Jean Berko (1958) set out to investigate what children learn about English morphology using nonsense words. Nonsense words were chosen since memorization alone could account for the fact that children use the correct inflection of actual words. However, if the child were to say that the plural of "wug" was "wugs", then the child must have internalized the regular rule for marking nouns as plural (this has since become known as the "wug test"). The subjects of this study included 4 and 5 year olds (preschoolers), 5 1/2 to 7 year olds (first graders) and adult college graduates. The subjects were shown cards with pictures and text. The text was read to the subjects and the subjects were required to fill in the missing word. Although Berko investigated plurals, possessives, adjectival inflections, derivations and compounds of nonsense words, as well as verb inflections, only her findings for past tense marking will be reported here, since this is the topic of interest. An example of her past tense verb marking materials was a card showing a picture of a man swinging an object. The text on the card was "This is a man who knows how to rick. He is ricking. He did the same thing yesterday. What did he do yesterday ? Yesterday he ________." Three other "regular" (regular in the sense that adults in the study chose to mark these words by adding the regular "ed" ending to indicate the past tense) nonsense words were used. Also included in the materials were the nonsense words "bing" and "gling", which were considered analogous to irregular verbs, one regular verb, "melt", and one irregular verb, "ring". The adult answers were considered to be correct (in most cases, the adult opinions of the correct inflection were unanimous) and the children were scored by comparing their responses to the adult responses. Two of Berko's findings are of interest here. First, for the irregular verb "ring", only 17% of the children provided the correct past form "rang", the rest of those who responded overregularized and said "ringed". Similarly, although 50% of the adults indicated that "bang" or "bung" was the past tense form of "bing" and 75% of the adults said "glang" or "glung" was the past of "gling", only one of the 86 children said "glang", the rest of those who responded again appeared to overregularize and supplied "glinged" and "binged" as the past tense form of "gling" and "bling" respectively. Taken along with her results for other types of inflectional and compound markings, Berko spoke of a language development system which entails consistency, regularity and simplicity, and models the marking of new words on frequently used regular morphological patterns. This process permits the child to be correct most of the time.

Susan Ervin (1964) was interested in studying the development of grammar in children. One of her areas of interest was past tense inflection. The data reported for this part of her study were culled from frequent texts of spontaneous speech collected from seven children ages 2 to 4. She noted that verbs with regular inflection were infrequent in the texts of the youngest children and therefore concluded that tense inflection begins with irregular forms. Interestingly, in some cases she found evidence of overregularization (e.g. "buyed", "comed") before the child had produced any other regular past tense forms. As for the developmental sequence of past tense marking, she reported that these children first learned the irregular forms as separate items of vocabulary (e.g., "did"). Next, for these highly frequent words, the children produced what she terms "analogic past tense forms", or what I have been referring to as overregularizations (e.g., "doed"). During this period the children oscillated between the correct irregular past tense form and the overregularization. She concluded that since these OR errors occurred before there was evidence of practice on the regular past tense forms, the tendency to regularize is very strong. These data, along with data dealing with pluralization, negation and transformations led her to theorize that language acquisition involves three processes. The child continually expands his ability to comprehend adult speech. However, since this practice from comprehension is not sufficient to explain the child's language production development, the child also uses imitation. Finally, children employ a process of building analogies of classes and rules. According to Ervin, this last process is the most significant in explaining language development. It is also this last process which would explain overregularization.

The spontaneous speech of three children, Adam, Eve and Sarah, was used by Courtney Cazden (1968) in order to study noun and verb inflections in language development. These data were collected as part of a five year study of language acquisition (Brown, Cazden & Bellugi, 1968). The data represent tape recorded spontaneous conversations between the children and their parents. In her analysis of errors in marking past tense forms, Cazden reported that the three children in the study often used the correct past form of irregular verbs before beginning to overgeneralize (which is equivalent to the term "overregularize" being used here). Specifically, 11 of the 32 irregular verbs which were overregularized had been correctly marked for the past tense by the children in an earlier conversation. For example, Eve used "came" correctly 11 times between the ages of 20 and 22 months, but between 25 and 27 months, she said "comed" three times. According to Cazden, this coexistence of the correct irregular past tense form and the overregularized form was common. This type of error is understandable, according to Cazden, because rote learning is required to learn both regular and irregular verbs, as well as the distinction between the two. Based on the fact that these children overregularized nouns when pluralizing by adding "s" to the irregular stem (e.g., "feets") and similarly marked the irregular verb stem using the regular rule (e.g., "tored"), she suggests that the irregular nouns and verbs may start out as separate lexical items. However, she does note that although this type of OR is fairly common in noun pluralization, accounting for 11 out of 32 overregularizations, it is relatively rare in past tense marking, where it accounts for only 5 out of 50 overregularizations. If irregular verbs do start out as separate lexical items, indistinguishable from regular verbs, the low frequency of this type of overregularization for verbs is surprising, considering the high frequency of irregular verbs.

In her overview of grammatical development theories, Susan Ervin-Tripp (1971) commented on Cazden's (1968) work. According to Ervin-Tripp, when English speaking children pluralize nouns and mark verbs for the past tense, there are three stages of development. First, children do not mark nouns and verbs at all. During the second stage, if the context of the conversation calls for marking, less than a third of the time, the child supplies a suffix. Presumably these two stages account for the initial period during which children correctly inflect both regular and irregular verbs, if they do so at all. The third stage, according to Ervin-Tripp, occurs when the child begins to overregularize, first on new or nonsense words, and then on irregular verbs. Ervin-Tripp accounts for this OR by postulating that the child generates the inflected verb through one of two routes. Either the child can use her "processing dictionary", presumably compiled through rote memorization, to locate a particular past tense verb, which may be stored as if it were a single, independent word, or the child can attach the regular "ed" suffix to the stem of the verb. These two routes account for the fact that children seem to freely alternate between the correct irregular past form (eg. "fell") and the overregularized form (eg. "falled"). It is important to note, however, that she makes no mention of exactly how a child makes the choice between these two routes.

Dan Slobin (1971a) sought to bring together researchers of various theoretical orientations and to have them comment on the same body of grammatical language development literature. The data were drawn from three of the aforementioned studies (Brown, Cazden & Bellugi, 1968; Cazden, 1968; Ervin, 1964) among many others. Overregularization was one of the topics of interest. Slobin reiterated claims made by both Cazden and Ervin, that children first use the past tense with irregular verbs, and they do so correctly (e.g., "fell", "went"). This is not surprising since these are the most frequent past tense forms used by adults. Then, as soon as the child learns perhaps as few as one or two regular past tense verbs (e.g.,"walked"), Slobin points out that the child "immediately replaces the correct irregular forms with their incorrect overgeneralizations from the regular forms". Noting that in some cases, as in those cited by Ervin, the overregularizations are seen even before the child produces any regular past tense forms, he says that it is apparently sufficient that a child only hears and understands the regular past tense in order to incorrectly apply it to irregular verbs. Slobin goes so far as to state that "even though the correct forms may have been practiced for several months, they are driven out of the child's speech by the overregularization, and may not return for years". Although extreme, Slobin's interpretation of the data was not uncommon at the time.

To this point this discussion has been confined to overregularization in English, but it has been observed in many languages, lending further support to the idea that this is a very general phenomenon. Slobin (1966), in his review of the Russian language acquisition literature, points out that Russian is particularly vulnerable to overregularizations since it is highly inflectional. For example, verbs are conjugated for person, number and gender of the subject noun in the past tense. Additionally, they are marked for three tenses. Simply stated, Russian morphology frequently uses suffixes of many types. It is no wonder then that overregularizations are widespread as the Russian child acquires his native language. Certain overregularizations in Russian show a developmental sequence analogous to the acquisition of the past tense in English. Zakharova (1958, cited in Slobin, 1966) found that as gender gains importance in classifying nouns, previously unused endings for each case begin to be used by the child. In a study of 200 children, Zakharova asked the children questions which required that they place both familiar and unfamiliar object names in another case form. She found that the youngest children disregarded the gender of the noun and used the suffix "om" as the universal instrumental and "u" as the universal accusative. These endings have a very high frequency in adult speech and are clearly marked acoustically. As gender gained importance in noun classification, the children began to use other endings, which are used to mark gender. It was only later that "om" and "u" re-emerged in the child's speech. Slobin points out that this parallels native English speaking children who initially use irregular verbs correctly in the past tense (e.g., "did"), then go through a period during which they overregularize (e.g., "doed") and finally return to the correct marking.

Popova (1958, cited in Slobin, 1966) also found evidence of overregularization among Russian children. She studied gender agreement between nouns and verbs in the past tense in children ranging in age from 1 year 10 months to 3 years 6 months. Popova told the children stories in the present tense and asked them questions about the stories in the past tense plural form, which does not require a gender distinction. In Russian, masculine nouns and their corresponding past tense verb forms are marked by a zero ending (they end in a consonant), while the feminine forms are marked by "a". The youngest children overregularized the feminine verb ending by frequently using it for both genders. The older children overregularized by frequently using the masculine zero ending for both genders. In yet older children, both forms appeared and were often used correctly. It appears as if two distinct stages of past tense overregularization occur in Russian. However, just as in English, OR is followed by a stage during which children improve their ability to correctly mark verbs in the past tense.

In 1962, Roger Brown, Ursula Bellugi, and Colin Fraser began a longitudinal study of English language acquisition at Harvard University (Brown, 1973). This study followed the progress of three preschool children, Adam, Eve, and Sarah. At least two hours of transcriptions of spontaneous speech between each child and his or her parents were obtained each month, over a period of 20 months for Sarah, and for almost six years for Adam and Eve. When the study began Eve was 18 months old and Sarah and Adam were 27 months old. Since children acquire language at widely different rates, the children were not equated by age, but instead by the length of their utterances, using both the mean length utterance (MLU) and their longest utterances (upper bound). Brown's analysis of past tense acquisition found that the children correctly used irregular past forms as early as Stage I (distinguished by a MLU of 1.75, and an upper bound of 5). In addition, Brown noted that irregulars were always more frequent than the regular past forms, as they also are in adults. Although infrequent, Brown recounts overregularizations in the speech of all three children, once again, during stages subsequent to those during which the children had correctly used the irregular forms. For example, during Stage IV (MLU = 3.5, upper bound = 11), Eve used the overregularized form "goed". However, within the same dialogue, she also said "went", adding further evidence to the contention that overregularizations freely alternate with the correct irregular form during the same developmental period.

Brown's (1973) investigation also included controlled elicitation experiments. A modification of Berko's (1958) "wug test" was employed. In Brown's experiments, the children, rather than the experimenter, supplied the context in the course of their spontaneous speech, therefore supplying a more representative sample of the full range of the child's usage than Berko's obligatory contexts supply. Although Brown supplies no specific examples of overregularizations in the marking of irregular past tense verbs, he does state that "children begin to overgeneralize inflections ... even before they attain our 90 percent criterion" (presumably this criterion refers to the percent correct for regular verb marking). So Adam, Eve and Sarah not only overregularize in the course of their spontaneous speech, they also appear to do so under more controlled elicitation situations.

According to Slobin (1973), the ability to construct a grammar, in any language, is essentially cognitive in nature. Therefore, language acquisition should be governed by a set of universal principles that are cognitive in nature. In accordance with this view, he stipulated "language-definitional universal principles", and based these principles upon the above mentioned English and Russian data, as well as data from thirty eight other languages. Slobin set forth "Operating Principle F: Avoid exceptions" to deal with the phenomenon of overregularization, about which he says "virtually every observer has noted some examples". "Universal F1" is a subdivision of Operating Principle F, which deals with the developmental sequence already noted. According to Slobin, linguistic marking of many semantic notions typically undergoes the following progression: 1) no marking, 2) appropriate marking in some instances, 3) OR (often accompanied by redundant marking), and 4) the full adult system. In some instances, various phases may occur during stage 3. As noted above, in Russian, one form of overregularization may drive out another (Zakharova, 1958; Popova, 1958, cited in Slobin, 1966). Since, according to Slobin, OR invariably occurs during the construction of the child's grammar, it seems a natural extension to postulate that overregularization itself is a basic cognitive phenomenon, not specific to language acquisition. Slobin argued that the development of specific language skills stems from general cognitive development. The child, possessing some inherent knowledge about the general structure and function of language, tries to express himself linguistically by attempting to understand the speech he hears. He analyzes linguistic input to ascertain meaning and in his pursuit, he is guided by his preliminary information about language, the processing constraints of his memory system and by general cognitive and perceptual strategies. Therefore as in all cognitive development, the acquisition of language relies on fitting new information to existing (and developing) structures, and the alteration of these structures to fit newly acquired information.

Stan Kuczaj (1977) analyzed spontaneous speech samples of 15 children in order to reaffirm some of the previously cited observations about the acquisition of the past tense in English. His subjects included 5 females and 9 males who formed a cross-sectional sample spanning ages 2 years 6 months to 5 years 6 months. The final subject, Abe (his oldest son), provided longitudinal information by participating in the study from age 2 years 4 months to age 5 years 1 month. As in Brown (1973), the children were not equated by age, but instead by MLU. Kuczaj's first analysis dealt with Brown's (1973) finding that children acquire the irregular past before they acquire the regular past form. Fourteen of Kuczaj's children reached 90% criterion for the regular past tense inflection during the study. At the time they reached this criterion all had MLU's above 3.0, and most of them had MLU's above 4.0, and all were approximately 3 years old. Therefore, just as in Brown's (1973) study, all were proficient with the regular past tense. However, only 3 of Kuczaj's children achieved 90% criterion for the irregular past during the course of the study. This sharply contrasts with Brown's (1973) results, which indicated that children acquire the irregular past before the regular past form. In fact, Abe reached 90% criterion for irregular forms 20 months after reaching criterion for regular verbs.

Kuczaj's (1977) next analysis dealt with overregularization errors. All of his children made OR errors, even if they had not achieved criterion on the regular past tense verbs. Kuczaj identified two specific types of overregularization errors. The children either attached the regular "ed" ending to an irregular stem (e.g., "eated"), or incorrectly inflected the irregular past form (e.g., "ated"). Kuczaj's results indicate that the earliest OR errors are of the "irregular stem + ed" form, while later errors tend to be of the "irregular past + ed" form. He points out that this early trend ("stem + ed" errors) was well supported by the data since the nine youngest children in the sample and Abe up until age 4 years made these errors, while the later trend ("past + ed" errors) was only a tendency, since only 4 of the 5 oldest children, and Abe in the last half of his fifth year, made these errors.

In order to lend further support to the idea that two distinct forms of overregularization errors appear in a specific developmental sequence, Kuczaj (1978) used an experimental grammaticality judgement task. In two experiments which employed somewhat different procedures, Kuczaj tested children from different age groups to see what past tense forms, including the two different types of overregularization errors, they judged as grammatical. Both experiments used samples of children who were divided into the following groups, based on age: Group I: 3-4 year olds, Group II: 5-6 year olds, and Group III: 7-8 year olds. In experiment 1, children were shown three puppets and told that sometimes the puppets "liked to talk silly". Four sentence pairs (for "nonchange" verbs whose past and present forms are identical) and ten sentence triplets (for "change" verbs whose past and present forms are different) served as stimuli, such that each sentence in either the pair or the triplet was identical except for the past tense verb form (e.g., "nonchange" pair: hit-hitted, "change" triplet: make-maded-maked). During each trial of the experiment, the child heard two or three of the puppets produce a different form of the sentence and was then asked if any of the puppets had "said something silly". The procedure used in experiment 2 was an extension of that used in the first experiment and included: 1) an elicited production task in which the experimenter's puppet supplied a future tense verb and the child's puppet had to respond in the past tense, 2) a judgement task which once again required the child to judge the "silliness" of a puppet's statements, and 3) two verb form choice tasks in which the child had to pick the past tense form she would use, as well as the form her mother would use. Ten "change" verbs were tested using this procedure.

The results from both these experiments support Kuczaj's (1977) previous findings. Group 1 children (3-4 years old) judged grammatical (e.g., "ate") and "base + ed" (e.g., "eated") overregularizations to be acceptable. Group 2 children (5-6 years) chose grammatical and "past + ed" (e.g., "ated") overregularizations as grammatical, and Group 3 children (7-8 years) rarely judged ungrammatical forms to be acceptable. Taken along with Kuczaj's (1977) spontaneous speech data, these data provide evidence for Kuczaj's theory that children go through a specific developmental pattern when learning the past tense of verbs. At first, children produce (or judge to be grammatical) some grammatically correct irregular past tense verbs. During this period they tend to avoid overregularizations. As the regular past tense inflection rule appears, children begin to overregularize. At this point they tend to avoid semantic redundancy and therefore when they do overregularize, they do so by attaching the regular "ed" suffix to the base form of the irregular verb. During the third stage of acquisition, children seem to be more driven by syntactic considerations, thus resulting in "past + ed" overregularizations which are semantically redundant, as well as "base + ed" forms. Gradually, with age, children learn that only one past tense form exists for each verb, and they avoid both types of OR errors. According to Kuczaj, this final stage probably comes about as a result of speech monitoring. Presumably the child keeps track of the fact that she rarely hears overregularizations in adult speech and therefore learns that these forms are unacceptable.

All of the investigations cited so far have implied that the correct form of irregular past tense verbs must be learned by rote memorization. Joan Bybee and Dan Slobin

(1982) investigated this assumption. They used both spontaneous speech records and Berko's (1958) "wug" test for real English verbs to study three age groups, preschooler's, 8-10 year olds, and adults. The irregular verbs they used were divided into eight categories : 1) no change verbs (e.g., "cut"), 2) "d" to "t" change verbs (e.g., "build-built"), 3) internal vowel change + "t" or "d" verbs (e.g., "feel-felt"), 4) vowel change, deletion of a final consonant + "t" verbs (e.g., "catch- caught"), 5) stem which ends in a dental + internal vowel change verbs (e.g., "ride-rode"), 6) internal vowel change of /i/ to /æ/ or to /ˆ/ verbs (e.g., "sing-sang"), 7) all other internal vowel change verbs (e.g., "give-gave"), and 8) stems that end in a dipthongal sequence + vowel change verbs (e.g., "fly-flew"). Based on the idea that frequency of exposure to individual verb forms is an important determinant of rote learning, Bybee and Slobin performed numerous correlations to ascertain if frequency actually determines the number of overregularizations. As expected, in many of these correlations, in all three age groups tested, they found a negative correlation between frequency and the percentage of OR's. In other words, the more frequent the past tense verb form, the less likely it was to be overregularized. Therefore, the analysis of their data indicated that, in general, irregular past tense forms are learned through rote memorization. However, as they point out, frequency alone cannot predict the error patterns for all of their data. They found evidence to indicate that at least some of these rote learned verbs are organized into classes based on the phonological characteristics of the past tense forms as indicated above. Bybee and Slobin use the term "schema" to describe these classes. When they compared the use of these schemas across age groups they found that all age groups use them, but that each age group used different ones, as evidenced by their OR errors. Preschool children overregularize by attaching the regular "ed" suffix , except when the verb already fits their schema for the past tense by ending in "t" or "d", in which case they do not mark it. Third grade children seem to have learned that you must mark a verb to indicate the past tense, and therefore they make more overregularization errors by adding "ed" to no change verbs than preschoolers do. Adults show evidence of having developed a set of schemas for vowel change verbs, and thus tend to make more OR errors based on these schemas.

B. Two Process Theories of Overregularization

1. Traditional Theory

As previously mentioned, the more traditional cognitive approach to the development of a theory to explain overregularization has been an interpretation which relies on two distinct cognitive processes. Most of the data from the investigations cited were originally explained by a theory which postulates the use of both rote memory and rule deployment. According to this theory, at first children use the past tense, if they do so at all, by utilizing rote memorization. They memorize the past tense forms one by one. If they hear their parents say "went" to indicate the past tense of "go", they will say "went" in the same context. During this early stage, they possess no past tense marking rules, and therefore they do not overregularize. As their exposure to the past tense increases, they begin to abstract the regular rule (add "ed" to the stem to form the past tense) from their parents' speech. It is at this point that the child starts to overregularize by over-applying this rule to irregular verbs.

As pointed out by Marcus et al., (1992), this theory suffers from a number of deficiencies. First, it doesn't correctly predict overregularizations in terms of a progression from rote only to rote + rule use. In its traditional form, it would predict that once the rote + rule stage is reached, the child should correctly inflect familiar irregulars and only overregularize novel irregulars. However, as already indicated, in numerous studies, in both spontaneous speech and in elicitation or judgement tasks, children do overregularize words that they had previously used in their correct past tense form.

Secondly, the traditional theory says little about the right hand arm of the U-shaped developmental sequence. Except for Kuczaj's (1978) supposition that somehow the child uses speech monitoring in order to learn that overregularizations are unacceptable, little is said about how children eventually stop making overregularization errors.

2. Marcus et al., (1992) Theory

Recently, a new theory has been proposed to better account for this phenomenon. However, like the traditional theory just outlined, this theory again relies on a two process system. Marcus et al., (1992) undertook an extensive analysis of the data dealing with the acquisition of the past tense. Their findings led to the formulation of a theory that fits the data as well as the traditional theory and answers some of the questions the traditional theory cannot. The theory is based on a blocking principle and memory retrieval failure. According to their Blocking Principle, an irregular form which exists in the adult's mental lexicon, blocks the application of the regular past tense marking rule (i.e., the presence of "broke" in the mental lexicon blocks the use of "breaked"). Based upon this principle, they propose that overregularizations in adults occur when the correct irregular form has not been heard before, and is therefore unavailable to block the application of the regular rule. Of course another possible explanation is that adults do not overregularize and say things like "breaked" and "wented" because they have never heard other adults say them. However, they point out that this would assume that adults would fail Berko's (1958) "wug" test for novel verbs, in which they would not have encountered either the "correct irregular" or the OR before, and as indicated in the previous discussion of Berko's results, adults clearly have no problem with the "wug" test and use both regular "ed" forms and forms analogous to overregularizations. Additionally, this alternate hypothesis would assume that adults would have problems creating or accepting regularized versions of the past tense of novel verbs (e.g., "scarf-scarfed"), and once again, the data show that they do not. Therefore, they proposed, the problem is not that the overregularizations have never been heard before, but that the correct irregular form has been heard and serves as a block to the OR. So the question now becomes, if adults use the blocking principle, why don't children ? Marcus et al., suggested that it is unlikely that children acquire the blocking principle from evidence about what forms are ungrammatical. Instead, it seems more probable that they determine which forms are ungrammatical using the blocking principle. They reached this conclusion by analyzing the feedback children receive from their parents when they produce ungrammatical utterances. Numerous studies, including Kuczaj (1977), have indicated that children do not receive negative feedback from their parents when they produce ungrammatical forms. However, if one presupposes blocking, the child would not require feedback that "comed" is incorrect; they would infer it from hearing "came". This blocking principle takes care of one of the traditional theory's problems; it explains how children might recover from overregularization errors (the right hand arm of the U-shaped developmental sequence), however, it does not explain why children would ever start overregularizing. It also does not explain why children have been shown to use both the correct and incorrect past tense form of a verb during the same developmental stage. To answer these questions Marcus et al., proposed that even though previously acquired irregular past tense forms may exist in the child's mental lexicon, in order for blocking to occur, these forms must be retrieved from memory. Since memory retrieval is imperfect and probabilistic, blocking may not occur and the child will overregularize.

According to Marcus et al., (1992), memory retrieval failures can also account for aspects of the phenomenon which are not handled by the traditional theory. First, the traditional theory does not address "past + ed" overregularizations such as "ated". According to the traditional theory, at first, children use rote memory to mark verbs for the past tense, then they begin to extract the regular rule for past tense inflection from their parents' speech and begin to overregularize. If it is assumed that some common past tense forms (such as "ate") are memorized in this early stage, before the rule is known, the traditional theory cannot explain why a child would then begin to overregularize the past tense form and use "ated" once he or she starts using the rule. Even if it is assumed that the child has not memorized "ate" in the early stage, the traditional theory offers no explanation for why a child would overregularize it in the later "rote + rule" stage. The memory retrieval part of the theory proposed by Marcus et al., can handle this problem nicely. Just as a failure to retrieve an irregular past tense form will disable the blocking mechanism and result in overregularization, failure to retrieve from memory a word's "past feature" (that is, when a word such as "ate" is retrieved, it is not recognized as a past tense form), will also disable the blocking mechanism and result in the OR "ated". Secondly, Marcus et al., in their seemingly exhaustive quantitative analysis of overregularization, have shown that different verbs have different OR rates. This issue is not addressed at all by traditional theorists. Marcus et al., however, offer a seemingly plausible explanation. According to their theory, overregularization results from failure to retrieve from memory an irregular form capable of blocking OR. Therefore, they hypothesize, it follows that forms with greater memory strength should be more resistant to overregularization. This hypothesis is borne out by two separate analyses.

First they examined frequency effects. They assumed that the more often a parent uses a particular past tense form, the stronger the child's memory trace for that form should be, and the stronger the association between it and the corresponding stem should be. Therefore, the frequency of irregular past tense forms in parents' speech should correlate negatively with the child's overregularization rate, and in fact they found that it does. The aggregate OR rates for 19 children negatively correlate with the aggregate parents' frequency counts. This is also supported by Bybee and Slobin (1982) who found a significant negative rank order correlation between children's overregularization rates and the frequencies of verbs in the speech of the children's caretakers.

Secondly, Marcus et al., (1992) examined the effects of "families" of similar irregulars. In this case they proposed that the higher the frequencies of the verbs in an irregular verb's family, the less likely that verb is to be overregularized. This idea was first introduced by Slobin (1971b) who suggested that "partial regularity" blocks overregularization. Slobin pointed out that many irregulars fall into groups which follow their own rules for forming the past tense (e.g., the irregular verb "break" falls into a category which also includes "come", "give" etc., that requires an internal vowel change in the formation of the past tense). These types of "partially regular" verbs seem to be more resistant to overregularization than irregular verbs which do not belong to a family of irregulars that bases the formation of its past tense on phonological similarities. This idea was also investigated by Kuczaj (1977) and Bybee and Slobin (1982). One result of these investigations was the finding that "no- change" irregular verbs (which includes verbs like "hit" and "cut" that require no change between the present and past tense) seem to be more resistant to overregularization than other irregular verbs. Bybee and Slobin (1982), as noted above, suggested that children form "schemas" for recognizing these patterns among irregular verbs. These "schemas" enable the child to more easily associate the past tense form with the verb stem and this presumably helps the child avoid overregularizations. The model proposed by Rumelhart and McClelland (1986), which is outlined below, lends further support to this idea. Marcus et al., set out to test this hypothesis, and once again, found that the data support the hypothesis. A negative correlation was obtained in 17 of 19 children for verbs with families whose stem and past form rhyme (e.g., the family for "stung" would include "clung", "flung", "swung"). Additionally, a negative correlation was found for all 19 children, for verbs whose families share a final consonant cluster and share the same change from stem to past form (e.g., the family for "stuck" would include "struck" and "snuck") and for verbs whose families share a final consonant and the same change from stem to past (e.g., this "family" for stuck" would include "stunk" and "slunk", in addition to "struck" and "snuck"). According to Marcus et al., these families contribute to the memory strength of the irregular verbs, and therefore their lower overregularization rates can easily be explained by the memory retrieval failure component of their theory. Traditionally, it was suggested that irregular verbs were stored as unstructured individual items. However, in light of the above discussion, this seems unlikely. Instead, Pinker and Prince (1988) suggested that perhaps the storage of irregular verbs also involves an associative network, based on the phonological similarities between the verbs. The analysis performed by Marcus et al., lends further support to this idea.

The overall picture which results from the theory proposed by Marcus et al., (1992) is as follows: 1) At first, possessing no rule for marking the past tense of regular verbs, the child only has rote memorization available to him. If he uses the past tense at all, he uses it correctly, based on what he has heard his parents say. This applies to both regular and irregular verbs. 2) When the regular rule is acquired, previously acquired irregular past tense forms can block overregularization, but they must be retrieved from memory and since this retrieval is imperfect, some OR occurs. 3) As the child develops, he hears the irregular form more often and strengthens this form in his memory, therefore improving his retrieval rate and allowing the blocking principle to work more and more effectively. As a result overregularizations gradually disappear. So, although this theory is similar to the traditional account in that it requires both the use of rote memory and a system of rules, in contrast to the traditional theory, there is no qualitative difference between children and adults, only a quantitative difference.

C. Non-Linguistic Evidence for Overregularization

1. Palermo and Eberhart (1968)

Thus far the discussion of overregularization has been limited to linguistic evidence. However, if this is truly a general cognitive phenomenon, there should be non-linguistic data supporting it. One body of non-linguistic evidence comes from a series of experiments conducted by David Palermo and his colleagues in the late 1960's and early 1970's (Palermo & Eberhart, 1968; Palermo & Howe, 1970). Using a modification of an experimental paradigm developed by Esper (1925), Palermo and Eberhart's experiment III set up a situation which they believed to be analogous to the problem of language acquisition which children face when beginning to learn the past tense of English verbs. They used non-linguistic stimuli and college students as subjects. The stimuli in the experiment were two-digit numbers and the responses were two-letter bigrams. A study - test procedure was employed. The study phase of the experiment consisted of the presentation of 12 "regular" pairs. These pairs were considered regular in that the response followed a number of specific (unstated) rules based on the stimulus. There were 16 possible regular pairs based on a 4x4 matrix. This matrix consisted of four digits, "6","7","8", and "9", which could occupy the first position in the stimulus, and four digits, "1","2","3", and "4", which could occupy the second position in the two digit stimulus. The regular rule specified that, the first position stimulus digits "6","7","8" and "9" required responses of "V", "H", "R" and "X" respectively, and the second position stimulus digits, "1", "2", "3" and "4" required responses of "M", "F", "G" and "K" respectively. Therefore, for example, the stimulus "61" would require a regular response of "VM", the stimulus, "83" would require a regular response of "RG", etc. During the study phase, two of the possible 16 regular pairs were omitted. Additionally, there were two presentations each of two "irregular" pairs. These pairs were irregular in that the responses required did not follow the above mentioned rules. So, for example, instead of the presentation of the regular pair "61-VM", a subject might be presented with the irregular pair "61-DL". The regular pairs were considered analogous to verbs which form their past tense using the regular "stem + ed" rule, the omissions were analogous to previously unencountered verbs, and the irregular pairs were analogous to verbs which do not follow the regular rule for past tense inflection. Following each study phase, the subjects participated in a test trial in which they were presented with the 12 regular stimuli, the 2 omitted regular stimuli, and the 2 irregular stimuli, one at a time, and were required to supply the response. No feedback was given during this test phase. A criterion of 3 successive errorless trials was set, and any subject that did not reach this criterion by trial 25 was dropped from the study. Results indicated that the irregular pairs, which were each presented two times per study trial, were learned first, followed by the acquisition of the regular pairs, which were only presented one time each, even though the total number of different regular pairs far outnumbered the total number of different irregular pairs. This is analogous to the language acquisition data which indicated that children learn the past tense forms of irregular verbs first (they are among the most frequently used verbs in English) and then learn the past tense forms of regular verbs, which are less frequently used, but more numerous. More interestingly, to this discussion at least, overregularizations of the irregular pairs were noted. Of the 69 errors made after a criterion of one errorless trial, 46 were overregularizations. Twelve of these were partial overregularizations, in that only one of the letters in the response followed the regular rule, and 34 were complete overregularizations. It was also pointed out that in many cases these ORs occurred well before criterion was reached on the regular pairs. This was considered comparable to Ervin's (1964) observation that in some cases children overregularize even before they show evidence of having mastered the regular past tense inflection rule.

2. Palermo and Howe (1970)

In an attempt to improve upon their past tense acquisition analogy, Palermo and Howe (1970) modified the stimuli used in the experiment by Palermo and Eberhart (1968). Since verb inflection is contingent on the preceding phoneme, but not the entire morpheme, and since there are actually three types of regular phonological past tense suffixes (/t/, /d/, and /Id/), three different single letter responses were required in response to the presentation of two digit numbers. For the "regular" pairs, the letter response was independent of the first digit, but was contingent upon the environment specified by the second digit. For example, the letter "F" was considered the correct response for all two digit stimuli with "7", "5" or "9" as their second digit. Additionally, "irregular" pairs, which required attending to both digits of the stimulus, were used (e.g., the stimulus- response pair 85-U). Since children acquire language from spoken input, the stimuli were auditory and the responses were made orally. The study-test procedure was again used with each study trial consisting of 12 regular pairs and 10 irregular pairs. One of the irregular pairs was presented four times, another three times, a third twice and the final irregular pair was presented once in the study session. All subjects participated for a minimum of 25 trials with a criterion of three errorless trials, or for a maximum of 35 trials. Subjects who did not reach criterion by this point were discarded. The results indicated once again that irregular pairs were acquired first, by rote memorization according to Palermo and Howe, and the acquisition of regular pairs (and presumably the regular rules) followed. The acquisition rate for the irregular pairs was positively related to the frequency of presentation. Overregularizations were again observed. All the OR errors totaled 158, and 64% of these occurred after at least one errorless trial for the irregulars but before criterion was achieved for the regulars. Once this criterion for the irregular pairs was achieved (and before criterion was reached for the regular pairs), there were a total of 188 errors, 101 of which were overregularizations. The number of OR errors was inversely related to the frequency of presentation of the irregulars. According to Palermo and Howe, these results suggest that the subject used at least two strategies in making her response. At first, rote memorization is used. Since the irregular pairs are presented more frequently, they are more easily memorized and therefore acquired first. Around the time the irregulars seem to be mastered, the rules governing the regulars seem to come into play. It is at this point that the subjects must decide which of the rote learned responses fit the rule, and which are the exceptions (irregulars). It is also at this point that most of the overregularization errors occur. This dual strategy hypothesis was supported by a further analysis indicating that initially, performance on regulars was around chance levels, but immediately following mastery of the irregulars, performance for regulars rose above chance.

The results obtained by Palermo and Eberhart (1968) and Palermo and Howe (1970) clearly indicate that the phenomenon of overregularization is not limited to language acquisition, and is not exclusively limited to learning in children. It suggests that OR is a more general cognitive phenomenon encountered when children or adults take on the acquisition of a rule based system. Although Palermo and his colleagues do not address the U-shaped developmental sequence of overregularization specifically, at least the left hand and the center of the U are suggested by their results which indicate that performance is initially high and then decreases as overregularizations become evident. This work will reanalyze Palermo and Howe's data, in order to examine the overregularization phenomenon in more detail.

Evidence will be sought for the right hand arm of the U-shaped curve, and for a situation analogous to the linguistic evidence for different acquisition rates for different verb stems. Palermo and his colleagues clearly subscribe to the traditional two stage "rote" and "rule + rote" acquisition theory, however, their results can also be explained just as well by the blocking and memory retrieval failure theory proposed by Marcus et al., (1992).

3. Others

Annette Karmiloff-Smith and Barbel Inhelder (1975/76) offer further evidence of overregularization in a non-linguistic setting. They performed a series of experiments which analyzed action sequences of children between the ages of 4 and 10 years old in a block balancing task. This was more of an attempt to understand general cognitive processes than a specific examination of block balancing. Various stages were observed in this task. First the children placed the blocks at any point of contact, let go, and tried again if the block did not balance. Next the children placed the block at any point of contact, "pushed hard" at the contact point, and then let go. The third stage of development involved a detailed examination of the block to be balanced and attempts to balance the block based on a number of different dimensions (lengthwise, widthwise, upended, etc.). The youngest children who reached this stage continued this "discovery of the block's properties" even if they balanced the block. Older children entered a fourth stage during which balancing the block resulted in the children no longer examining the block. During the fifth stage, the children began to search for a point of balance for each block. This is the stage during which a "theory in action" became apparent. The children would first try the geometric center of the block as the point of balance, and since this worked for a number of blocks, they tended to try this for all blocks, even those that had been previously balanced at a different point. Therefore, it is during this stage that U-shaped learning and overregularization are evident for those blocks that do not follow the geometric center balancing point theory in action. Evidence of U-shaped learning and OR led Karmiloff-Smith and Inhelder to point out that their "results seem to suggest certain functional analogies between the acquisition of physical knowledge and the acquisition of language."

Non-linguistic evidence for U-shaped development has also been presented by Sidney Strauss (1982). In a study of children ages 3 through 11, he examined the children's ability to solve problems about the magnitude of temperature. The children were presented with three cups containing equal amounts of the same temperature cold water (A,B and C). Water from cups A and B were poured into an empty cup D and the children were asked to compare the temperature of the water in cups C and D. He found a U-shaped behavioral growth curve in which the youngest children (ages 3-5) correctly recognize that the cups contain the water of the same temperature, children ages 6-9 often incorrectly state that the two cups contain water of differing temperature, and finally, the oldest children, ages 10-11, once again correctly respond to the task.

D. One Process Theories of Overregularization

1. Rumelhart and McClelland (1986) PDP Model

a. Description

As previously mentioned, there are two fundamentally different schools of thought regarding an explanation for overregularization. Thus far, the two process view, that the phenomenon can be best explained using two separate cognitive processes (i.e., rote memorization and the use of rules), has been detailed. The other view, that overregularization can be accounted for in a system without explicitly defined rules, stems from the work of Rumelhart and McClelland (1986). Their work represents a significant departure from most traditional cognitive approaches and although this discussion will be confined to overregularization, it addresses many perceptual and psychological phenomena. They have developed a connectionist model of the human mind called Parallel Distributed Processing (PDP), which takes a bottom up approach in that it strives to pattern the workings of the mind on neuronal activity. Although the specifics of the system can be rather complicated, the basics of the specific model they developed to handle the phenomenon of overregularization will be outlined. How well the system mimics the language acquisition data discussed above will then be addressed, with an emphasis on the differences between the theory suggested by their model and the more traditional two system cognitive theories which deal with overregularization.

First it should be pointed out that the model developed by Rumelhart and McClelland (1986) is a computational model. However, unlike most computer implementations which rely on serial processing, it utilizes parallel processing. Basically the model is a pattern associator network which uses a phonological representation of the verb stem as its input and produces a phonological representation of the past tense form of that stem as its output. The input and output units are linked by a pattern associator that consists of modifiable weighted connections. The structure of the model is divided into three fundamental parts: a fixed encoding network for the input (the phonological representation of the verb stem), the pattern associator which learns the relationship between the input and the output, and a decoding network which produces the output (a phonological representation of the past tense form of the verb). Each verb stem is represented by a set of nodes representing the specific phonological components of that verb. When this set of input nodes is activated, each node sends output, equal to its activation level multiplied by its link weight in the pattern associator[1], to the output nodes that it is connected to. Each output node then sums its weighted inputs and compares the result to a threshold. If the threshold of each output node is exceeded, the unit is turned on and the set of activated output nodes comprise the phonological representation of the past tense verb. Actually a decoding network is used for converting a featural representation of the past tense form, which may not be an exact match, into a legitimate phonological representation.

The model "learns" when a "teacher" supplies the network with pairs of patterns consisting of a representation of the verb stem and a representation of its corresponding correct past tense form. Using the perceptron convergence procedure, the network compares its version of the past tense form with the correct version. This comparison results in the adjustment of the connection weights in the pattern associator between the input and output units. At this point the thresholds for each output node are also adjusted. In this way, the model improves its performance and is able to generalize to novel inputs, based on their overlap with forms it has already encountered.

---

[1] Initially all the connection weights in the pattern associator are set to zero.

There are no specific representations of rules, and no distinction is made between "regular" and "irregular" verbs.

Prior to their large scale implementation of a network based on the phonological representations of verb forms, Rumelhart and McClelland (1986) set up a simpler network which can be thought of as analogous to past tense acquisition in a manner similar to the Palermo experiments (Palermo & Eberhart, 1968; Palermo & Howe, 1970). Their "rule of 78" used representations of inputs and outputs consisting of three digit numbers. The digits 1,2 or 3 could occupy the first position, the digits 4, 5 or 6 could occupy the second position, and the digits 7 or 8 could occupy the third position of the number. The rule was simple; If the input number had an 8 as its last digit, the output number would consist of the same initial two digits, with 7 as its last digit, for example (258) -> (257) and if the input had a 7 as its last digit, the output would consist of the same initial digits, with 8 as its last digit, for example (257) -> (258). This rule is meant to be analogous to the regular "stem + ed" rule for inflecting verbs in the past tense. As an analogy to irregular verbs, an exception to the "rule of 78", (147) -> (147) was used. Initially the network was exposed to 20 presentations each of one example of a regular pair and the exception. The model did not achieve perfect performance at this point, but it came close (90% correct for both the pair which followed the rule and the exception). At this point it is irrelevant to the model that one of the patterns is an example of the rule, and the other is an exception. According to Rumelhart and McClelland, this is similar to the situation encountered by children as they initially begin to learn the past tense when exposed to only a few high frequency verbs. The model is then presented with all 18 patterns, only one of which is the exception. After 10 presentations of all 18 patterns, the model began to learn the set of connection strengths that represents the regular rule, however, at this point it was not performing well on the exception, but was overregularizing it (i.e., (147 -> 148). It is only at the point that very few errors are made on the 17 regular patterns, well after 40 exposures to all 18 patterns, that, as Rumelhart and McClelland state, "it begins to accommodate the exception". By 500 presentations of all patterns, the model is sufficiently competent on both the regular patterns and the exception to correctly respond to both types of patterns consistently. This is roughly analogous to a child's acquisition of the past tense. At first, with only a very limited exposure to the past tense, children perform well on the limited number of both regular and irregular verbs which they use. As they encounter more examples they begin to overregularize irregular verbs and their overall performance decreases. After a significant amount of exposure, their performance increases slowly for both regular and irregular forms and overregularizations decrease. So, in a system without explicitly defined rules or a qualitative distinction between irregular and regular forms, Rumelhart and McClelland demonstrated overregularization and the U-shaped acquisition curve.

In order to develop a system which more closely resembles past tense acquisition, Rumelhart and McClelland (1986) devised a PDP network containing two sets of 460 input units which represented the different phonological components of the base form of a verb, and 460 output units representing the phonological components of the past tense verb forms. The system employed "Wickelfeatures", based on a scheme proposed by Wickelgren (1969). In order to obtain the U-shaped developmental sequence observed in the language acquisition literature in one network, Rumelhart and McClelland made a number of assumptions. First, they assumed that the verbs with the highest frequency are irregular. This was supported by the fact that the ten verbs with the highest frequency in the word frequency count made by Kucera and Francis (1967) are all irregular verbs. They then assumed that children acquire verbs in order of decreasing frequency. Therefore, initially they learn mostly irregular verbs, and it is only later on that they encounter more and more regular verbs. Most importantly, as pointed out by Marcus et al., (1992), they assumed that at some point children show an "explosive" vocabulary spurt, resulting from a sudden influx of regular verbs. In their model, Rumelhart and McClelland assumed this vocabulary spurt occurs after the child has just acquired his tenth verb. This assumption translated into a network which was trained in two phases. At first the network was presented ten times each, with the 10 highest frequency verbs (excluding "do" and "be" which can also be auxiliaries), of which only 2 are regular. The model was then presented 190 times each, with this initial list of verbs, plus the 410 next most frequent verbs, of which 80% are regular. During the initial phase of training, the model successfully learned the past tense of the 10 verb stems which it was presented. As the second phase began, the model began to overregularize the irregulars, and then, over the course of this second phase of training, the network's performance improved and overregularizations decreased until its overall performance asymptoted at approximately 90% correct by the 200th epoch. Once again, in this detailed network, which presumably resembles more closely the features of the English language, Rumelhart and McClelland demonstrated that the overregularization phenomenon and its U-shaped developmental sequence can be observed.

b. Criticisms

As with any dramatically new approach to a well-established problem, Rumelhart and McClelland's (R&M) connectionist model of past tense verb learning has met with considerable criticism. An entire issue of Cognition (1988, volume 28) was devoted to a discussion of the issues and theoretical implications raised by R&M's model. In essence, R&M's simulations sparked a heated theoretical debate, which continues between the one and two process language acquisition (and general cognition) camps to this day.

Some of R&M's critics, including Fodor and Pylyshyn (1988), challenge the theoretical implications of the PDP model. Fodor and Pylyshyn (F&P) claim that cognitive processes, including past tense acquisition, require representations which are systematic (i.e., representations which have structure). According to F&P, a connectionist architecture, like that proposed by R&M's PDP model, lacks hierarchical structure since it handles mental representations as lists. In the domain of language acquisition, F&P argue that these mental lists do not take into account the syntactic or semantic structure that is inherent in language. As such, F&P claim that connectionist (or one process) theories do not sufficiently describe a general cognitive architecture. In other words, according to F&P, a PDP model should not be considered a theoretical model of cognition, but may instead be viewed as a possible model of implementation.

Other criticisms, most notably Pinker and Prince (1988) and Lachter and Bever (1988), target the specifics of the R&M implementation of past tense acquisition. Both Pinker and Prince (P&P) and Lachter and Bever (L&B) point out numerous problems with the PDP model, but their basic criticisms can be summarized as follows:

1) According to P&P and L&B, the Wickelfeature representation used in the R&M network poses many problems. However, since these problems have been generally been recognized (as evidenced by the fact that all other subsequent connectionist models use a different representational scheme), and since the SDM simulations performed as part of this work do not use this type of representation, the specific criticisms are not relevant here.

2) The U-shaped learning claimed by R&M is actually the result of a discontinuous input set. Since R&M initially only train the model on 10 high frequency verbs (8 of which are irregular), they artificially induce the initial high performance rate. Overregularization, and a decrease in performance on the irregulars, is only seen during R&M's second phase of training, which overwhelms the system with regular verbs.

Although valid criticisms, the criticisms leveled at the R&M model which deal with the Wickelfeature representation of items, do not necessarily discredit the model itself. Had a different representational design, which better handles the characteristics of the English past tense, been implemented using the same architecture, it is entirely possible that U-shaped learning would have occurred. However, P&P and L&B's second criticism is much more damaging to R&M's claims. There is little behavioral evidence that children undergo any type of explosive vocabulary growth at the time they begin to overregularize (Pinker, 1991). R&M's training schedule, which cannot be considered comparable to a child's learning experience, may have been solely responsible for the U-shaped learning the R&M model exhibited.

2. MacWhinney and Leinbach (1991) Model

A number of researchers have tried to address the criticisms of the R&M model in different connectionist implementations of past tense acquisition. In 1991, MacWhinney and Leinbach (M&L) claimed to achieve U-shaped learning in a network model which differed from the R&M model in its phonological representation, architecture, learning algorithm and input corpus. M&L's representational scheme accounts for 8 different vowel features, 10 different consonant features and the position of the phoneme within the syllable. M&L also dedicate five units to the most common verb tenses (present, past, present participle, past participle and 3rd person singular present). Unlike R&M, M&L utilized a multilayered network architecture which included two layers of 200 hidden units each. In addition, M&L supplemented this network with a set of direct connections between the input and output nodes which copy the phonological "left justified" form of the present tense directly onto the output. Instead of the perceptron convergence learning algorithm employed in the R&M model, M&L used back-propagation, which is considered a more powerful learning algorithm. M&L's input corpus included 5481 forms, including 118 past tense forms, and was therefore more comprehensive than R&M's. Training of the M&L model was based on the relative frequencies of the forms based on Francis and Kucera (1982) so that the most frequent forms were trained once per epoch, and the rarest forms were presented once every 700 epochs. A total of 24000 epochs were run, resulting in the presentation of approximately 1.3 million forms. Finally, M&L's training schedule was continuous, as opposed to the discontinuous training used by R&M.

M&L claim to produce three of the four components of U-shaped learning, including overregularization, coexistence of correct and incorrect forms and final correct usage. Although regular forms in the M&L model exhibit early correct usage, early correct usage for the irregulars was not seen (except for verbs like "hit" which are not marked in the past tense). During the initial learning stage in this model, most of the irregulars produced were either unmarked or overregularizations. Although it is difficult to evaluate the validity of these claims since M&L present no graphical or statistical analysis of their data, one point should be made. Critics claim that R&M achieved early correct usage through a discontinuous training set. M&L's training set was continuous, but it did not produce early correct usage. So the question remains as to whether a discontinuous training set is a necessary requirement for U-shaped learning in connectionist architectures. In addition, Marcus et al., (1992) point out that M&L, because of the direct connection between input and output nodes which essentially copies the input to the output, have designed a model with two innate pathways, one for the regulars (the direct input-output connection) and one for irregulars (through the hidden units of the multilayered network). This is reminiscent of the L&B claim that the Wickelfeature representation used by R&M in some ways embodied past tense acquisition rules. If these claims are valid, both systems (R&M and M&L) in some sense use two processes, rather than the one process implied by a purely connectionist model.

3. Plunkett and Marchman (1991,1993) Models

Plunkett and Marchman (1991) (P&M-91) presented a network model which also addresses some of the R&M criticisms. P&M-91's phonological representation differed from both R&M and M&L. They constructed an artificial language consisting of randomly generated, possible English CVC, VCC and CCV strings. Each consonant and vowel was represented by six feature units (e.g. voiced/unvoiced;

front/central/back; consonant/vowel). Irregulars were assigned to one of three categories: 1) identity mapping (hit-hit), 2) vowel change (come-came) or 3) arbitrary (go-went). P&M-91 used a multilayered network with one layer of 20 hidden units and utilized the back-propagation learning algorithm. Although they varied a number of parameters in different simulations, the ones of greatest interest here (because they most closely approximate both the input and the results of English speaking children) are the "phone" simulations. In these simulations P&M-91 presented the network with a corpus consisting of 410 regulars and 2 arbitrary, 20 identity and 68 vowel change irregulars. In addition, to some extent they mimicked the phonological characteristics of irregular verbs. All identity irregulars were required to end in a dental, all vowel change stems were restricted to 11 possible English VC endings, and they did not use this phonological information to define the irregulars (that is, some regulars also possessed these properties). As in M&L, a continuous input set was used.

P&M-91 argue that there are actually two types of U-shaped learning, macro and micro U-shaped learning. Both are characterized by initial correct usage, for both the irregulars and the regulars, and recovery from OR for the irregulars. The difference between the two lies in the onset and extent of overregularization observed. In macro development, OR onset is fairly abrupt, it is applied as soon as the regular rule is abstracted. In addition, macro development, according to P&M-91, is comprehensive, as OR's occur for all irregulars during this period. In contrast, micro U-shaped development is distinguished by a gradual and non-absolute onset of OR. Micro development is non-absolute in that OR is selective, only some verbs are overregularized; it is gradual since each irregular verb starts overregularizing at a different point in time. According to P&M-91, it is actually micro U-shaped development which is observed in the language acquisition literature. P&M-91 present results for their "phone-34" simulation which shows micro U-shaped learning for all three types of irregulars. In other words, the past tense of some stems, in all three classes of irregulars, are correctly produced at some point, subsequently overregularized, and then finally produced correctly again. However, initial correct performance on the irregulars was not achieved in this simulation (just as it was not achieved in M&L). Marcus et al., (1992) have argued that lacking this initial period of correct performance (which is supported by behavioral data), P&M-91's micro U-shaped development can be described as "any wiggle in a developmental curve". P&M-91 make a good point in defense of this shortcoming. Prior to beginning acquisition of the past tense, children have acquired considerable information about phonology. The simulations must acquire this phonological information <u>at the same time</u> that they are acquiring the past tense. In addition, P&M-91 claim that children are not required to learn a large number of verbs initially, as are their simulations, and therefore it is understandable that the model is not perfect initially. However, although this may explain why the connectionist models of M&L and P&M-91 do not show early correct performance, it does little to settle the issue of whether only a discontinuous input set, such as the one used by R&M, is capable of producing this aspect of U-shaped learning.

Based on parental report measures (Marchman & Bates, in press) that suggest that verb acquisition in children is a gradual process which proceeds incrementally, Plunkett and Marchman (1993) (P&M-93) implemented a connectionist model which used an incremental training regime. P&M-93's input corpus consisted of 458 regulars and 2 arbitrary, 20 identity and 20 vowel change irregulars[2]. Inclusion of an item in the vowel change class of irregulars was based on four allowable vowel-consonant clusters, but inclusion in the other two irregular classes was based on the same criteria as P&M-91. Unlike P&M-91, regular verbs were

---

[2] According to P&M-93, this distribution of items among the regulars and three types of irregulars is similar to the verb vocabulary of a child who has mastered the past tense of English in two ways: 1) the number of regulars greatly exceeds the number of irregulars and 2) arbitrary items are an order of magnitude lower than the other irregular items.

designed to more closely follow English morphology as follows: 1) If the stem ended in a dental consonant, then the past tense suffix was /-id/ (e.g., pat-patted), 2) if the stem ended in a voiced consonant or vowel then the suffix was voiced (e.g., dam-dammed), and 3) if the stem ending was unvoiced, then the suffix was unvoiced (e.g., pack-packed). Like R&M, P&M-93 initially presented the model with a small number of high frequency stems (10 regular and 2 arbitrary, 4 identity and 4 vowel change irregulars). During this initial training, the frequency of presentation per epoch for each type was 15 for the arbitrary irregulars, and 5 for the regular, identity and vowel change classes. This initial training continued until all verbs were correctly mapped to their past tense form. However, in contrast to R&M, P&M-93 then increased the size of the vocabulary gradually, one item at a time every five epochs until a vocabulary size of 100 was reached. Medium frequency items (token frequency per epoch = 3) were presented during these middle epochs, and then the lowest frequency items (token frequency per epoch = 1) were added at a rate of one per epoch.

Using the same criteria as Marcus et al (1992), P&M-93 defined the rate of overregularization as the number of OR tokens divided by the sum of the number of OR tokens and the correct irregular past tokens. They compared their simulation to Adam, one of the children in the Brown (1973) corpus, whose U-shaped development was plotted by Marcus et al (1992)[3]. Both Adam and the simulations shared the following characteristics: 1) a generally high level of performance on the irregulars (above 90%), 2) an initial period of perfect performance on the irregulars (if the point when the simulation had already mastered the initial 20 items is considered the start point for the simulations), and 3) a prolonged period (ages 35-55 months for Adam and vocabulary size 110-230 for the simulations) when low levels of overregularization are seen. This pattern is also similar to that reported by Marcus et al., (1992) for Eve and Sarah, two other children in the Brown (1973) corpus. In addition, the simulation returned to perfect performance on the irregulars when the vocabulary size increased beyond 230 items.

Although P&M-93 achieved U-shaped development which looks remarkably like the behavioral data presented by Marcus et al., (1992), they did so using a discontinuous input set, albeit a different type of discontinuity than that used by R&M. P&M-93 claim that their discontinuous training strategy is similar to parental report measures which indicate that verb acquisition is a gradual incremental process. According to P&M-93 and Marchman and Bates (in press) overregularization in children may arise from a process of mass action, that is, upon the achievement

---

[3] See P&M-93 Figure 3 a and b.

of a critical mass of regular verbs[4]. In the P&M-93 simulations, this occurred once the regular verbs accounted for more than 60% of the total vocabulary. Therefore, instead of requiring two processes as do traditional theories, their simulation suggests that U-shaped development may be the result of a one process system where overregularization is driven by gradual and incremental exposure to regular verbs.

4. Kanerva (1988) SDM Model

In 1988 Pentti Kanerva introduced a new theory called sparse distributed memory (SDM). This theory is a mathematical model of memory, inspired by the biology of the human nervous system. An investigation of this new theory is the primary objective of this work. Specifically, the goal of the SDM simulations described below (see Section II) is to determine SDM's ability to model overregularization and U-shaped learning.

Kanerva's (1988) SDM model is based on the idea that human long-term memory is a storage system which associates sensory input with behaviors and actions appropriate to the situation (Denning, 1989). Sensory input is represented in the model as long bit vectors which can contain thousands of bits. As it applies here, a bit is a binary unit that has two states, either ON, (represented by a "1") or OFF

---

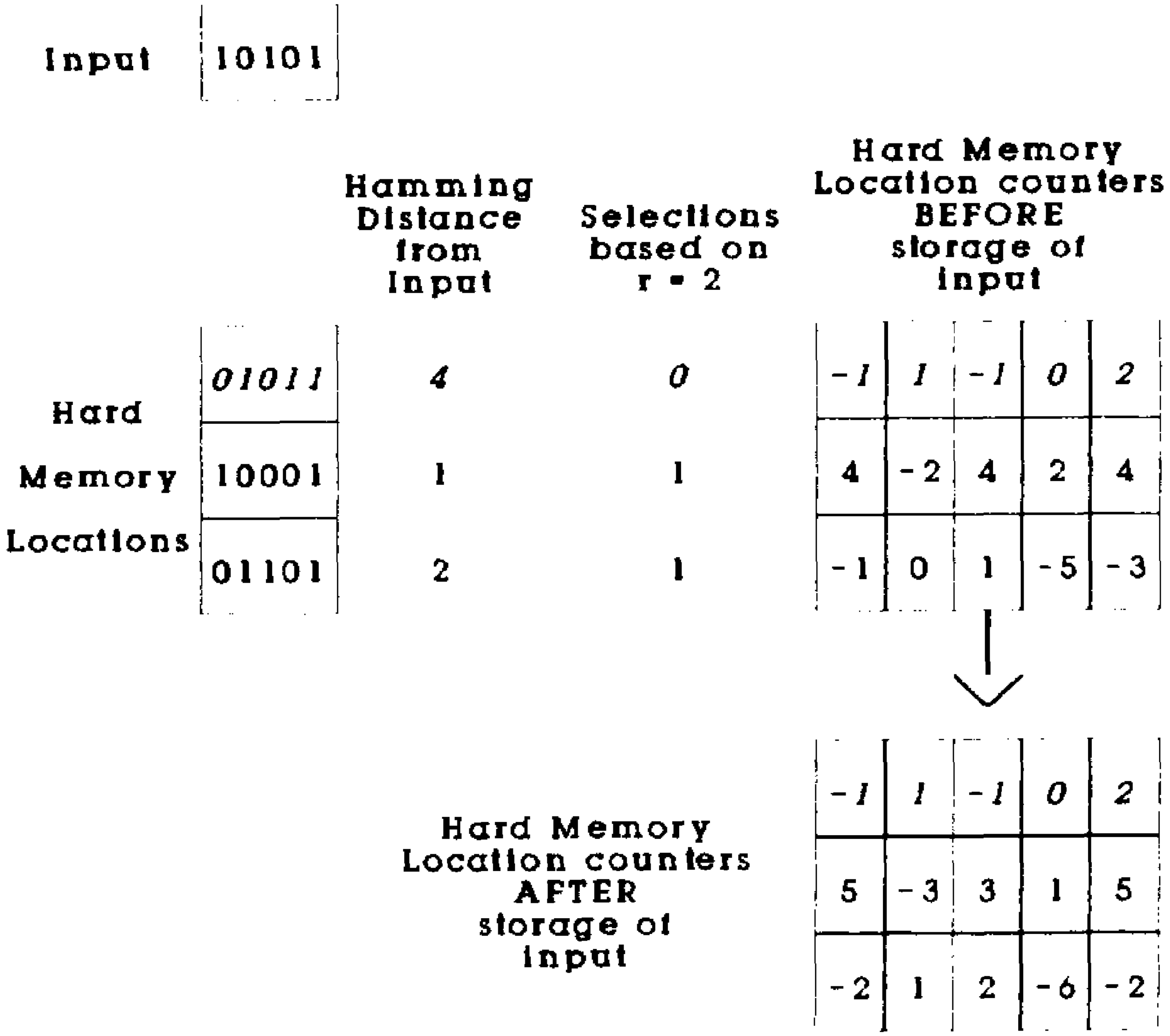
[4] Marchman and Bates (in press) provide behavioral evidence for this critical mass theory in children.

(represented by a "0"). So, each sensory input is represented in the system by a long string (vector) of "1"s and "0"s, for example "00101110...001". These input vectors enter a memory system which is composed of a very large number of addresses, which are also represented as long bit vectors. If each sensory input is represented by a 1000 bit vector (a string of 1000 "1"s and "0"s), as Kanerva suggests, then the total possible number of addresses is 2^{1000} . This is the size of the memory space. Kanerva points out that it is impossible that a memory space this large is actually used in the human brain, since the number of neurons in the brain is only approximately 2^{36}, and only 2^{100} molecules of water would be needed to completely fill the brain. He also points out that this number of memory locations is not even necessary, since it is also impossible that 2^{1000} separate entities would need to be stored in memory in a lifetime. Therefore, only a small number of the possible addresses in the total memory space need to be available. This subset of the total memory space, which is actually used for memory storage, is referred to as a set of "hard" memory locations. Kanerva proposes a system which uses 1,000,000 "hard" memory locations to store information which he represents as 1000 bit vectors. Since only 1,000,000 locations of the total memory space, which encompasses 2^{1000} locations, would actually be used by this system, the system is a sparse representation of the total memory space. Each storage location, whose address is represented by a 1000 bit vector, can be thought of as analogous to a neuron which has fixed input synapses and modifiable output synapses.

Each input enters the system via an input register for the data pattern, with each register holding the 1000 bits of the input. This input register is then compared to all the "hard" memory locations in the system. Each location in the system has an address decoder that compares its own 1000 bit address with the input. If the distance between the input and the address of the location is within a specified radius, r, the location selects itself and the input is stored there. The distance between the input and the address of the location is actually the "Hamming" distance, which is a measure of the dissimilarity between the patterns of "0"s and "1"s. For example, the Hamming distance between the bit vectors "0011001" and "1010101" is 3 bits, since the bits in positions 1, 4, and 5 are different in the two vectors. The radius suggested by Kanerva for his system, which uses 1000-bit input vectors and one million 1000-bit memory locations, is 451, since 1/1000 of the addresses are within 451 bits of any given address. Therefore, when an input enters this system, all locations whose addresses are within a Hamming distance ot 451 bits of the input will select themselves, and the input will be stored at these locations. It is for this reason that the memory is considered distributed, since each input is stored in many locations, rather than just one.

Each input will be stored at many locations within the specified radius, and every location will store information from many different inputs. Therefore, it is important that a new input does not simply replace the information already stored in the selected location. What happens instead is that each location is actually a set of counters. When an input is selected for storage by a location, a "1" is added to every counter in the memory location corresponding to a "1" in the input at that position. For each "0" in the input, a "1" is subtracted in the corresponding memory location counter. For example, assume that the location whose address is "10001" is selected for storing the input "10101". Suppose that at the time of selection, the counters for the location "10001" hold "4,-2,4,2,4". After the input "10101" is stored at location "10001", its counters would hold "5,-3,3,1,5" (see Fig. 1).

In order to retrieve a 1000 bit pattern corresponding to an input, the system again selects the memory locations within the specified radius based on the Hamming distance between the input and the memory locations, just as it does for storage. The system then mathematically constructs the output by adding the counters for each position in all the selected locations. If the result of this addition is negative or zero, the output for that position is "0", and



Hard Memory Locations	Hamming Distance from Input	Selections based on r = 2
01011	4	0
10001	1	1
01101	2	1

Hard Memory Location counters BEFORE storage of input

-1	1	-1	0	2
4	-2	4	2	4
-1	0	1	-5	-3

Hard Memory Location counters AFTER storage of input

-1	1	-1	0	2
5	-3	3	1	5
-2	1	2	-6	-2

Figure 1. Storage in SDM.

if the result is positive, the output position is "1". For the sake of simplicity, in both of the examples which follow, SDM is implemented in an autoassociative manner, that is, the input is stored in a location whose address matches the input pattern. As will become evident later (in the SDM simulations described in section II), this is only one way in which SDM can be implemented to store and retrieve patterns. Assume an address of "10101". Also assume that only the following 3 locations are selected, "00101", "10100" and "10001" (which, for simplicity's sake, are all within a Hamming distance of 1 bit from the input). Assume, once again that the counters for these three locations, at the time of selection, hold "1,-2,0,-4,0", "-1,-1,1,-3,-5" and "3,-4,2,1,6" respectively. The resulting addition of the contents of the counter for each position in these three locations would be "3,-7,3,-6,1", which would result in an output of "10101", a match (see Fig. 2). Figure 3 shows a similar situation, but this time the counters have different values, which results in an output pattern which is a partial match, rather than an exact match, to the input pattern.

The distribution of inputs among many locations not only allows retrieval of partial matches, it also makes for a robust system which will not completely lose stored information if one or more of the locations are destroyed. According to Kanerva (1988) this is analogous to the non-

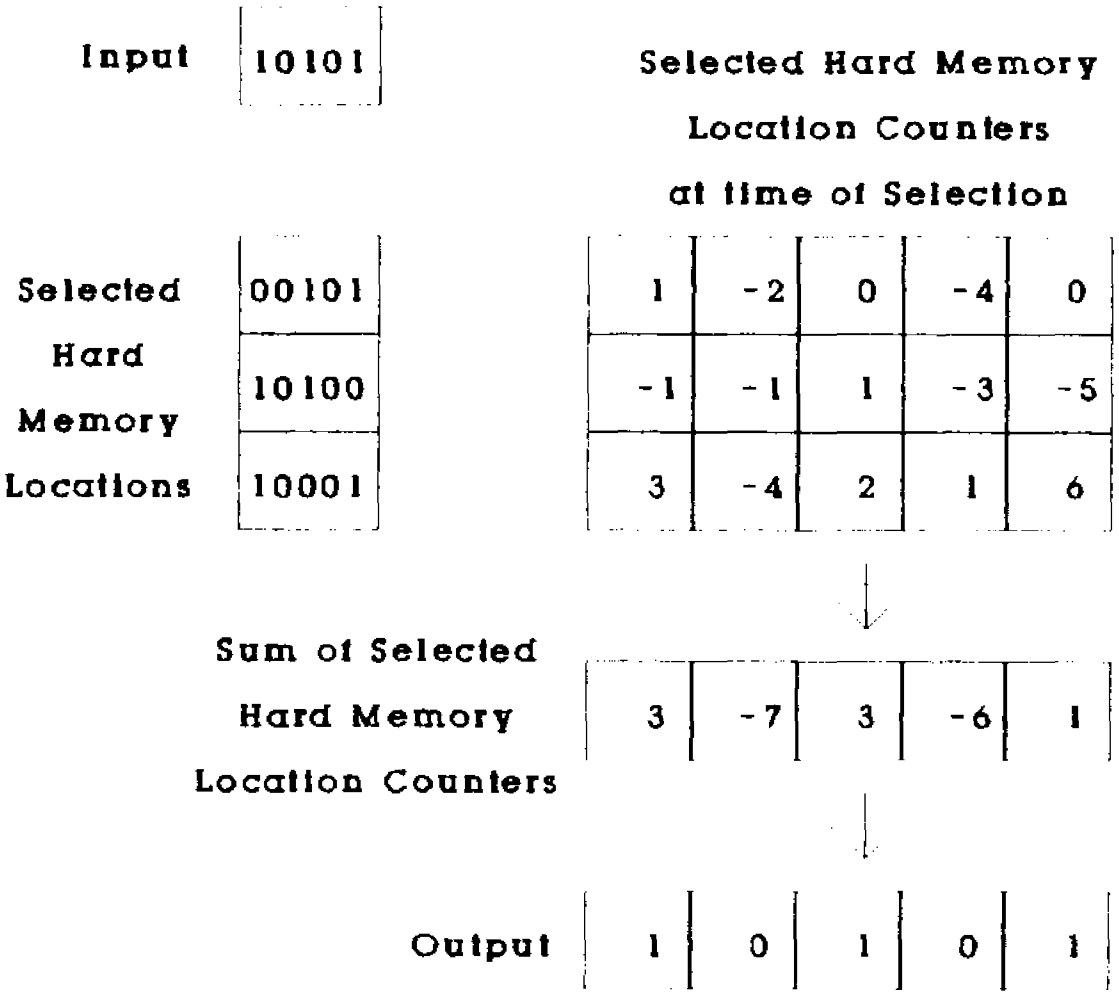


Figure 2. Retrieval from SDM - An exact match.

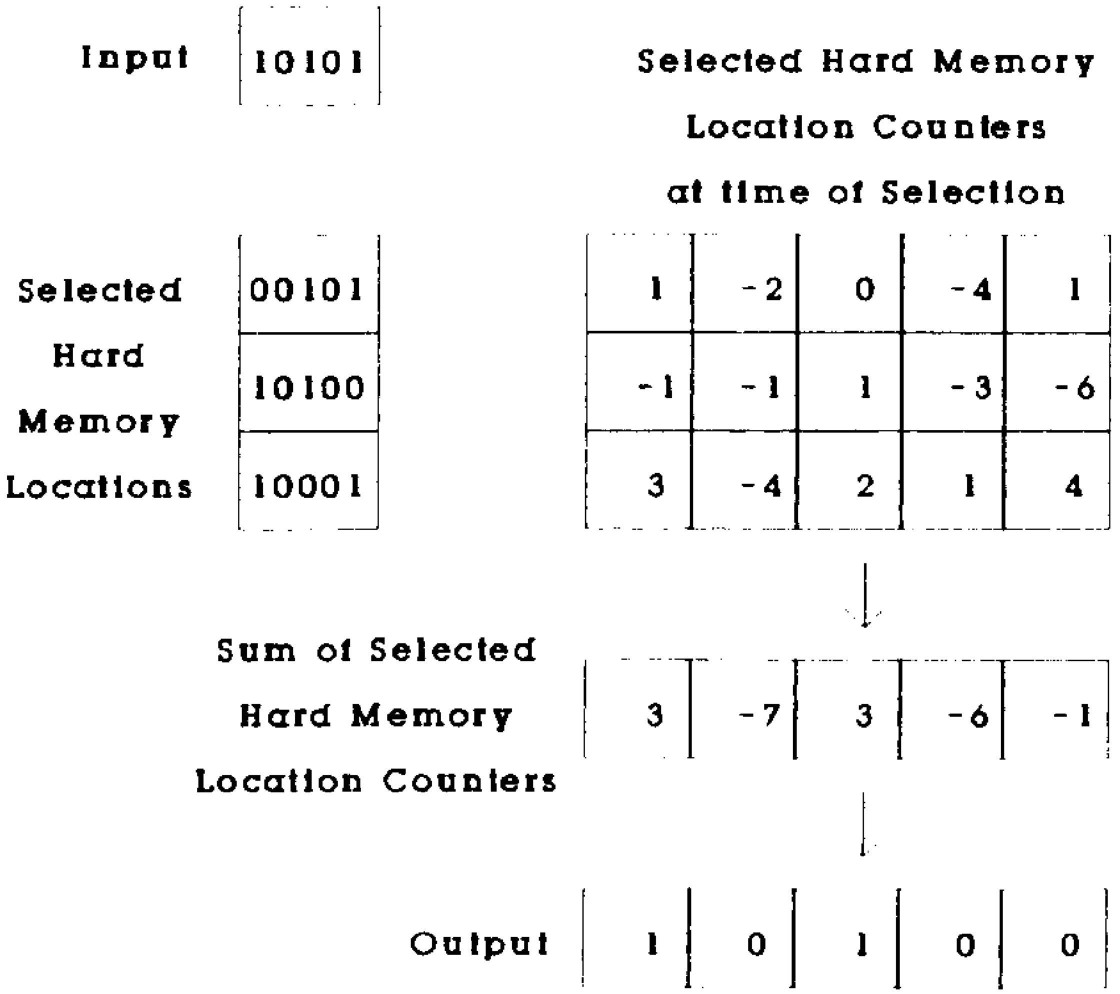


Figure 3. Retrieval from SDM - A partial match.

specific effects of brain damage. If it is assumed that the brain also stores information in many locations, as the sparse distributed memory system does, then it is unlikely that localized brain damage would destroy all copies of the stored memories. Instead, only some copies would be destroyed, resulting in less effective retrieval, rather than total failure.

5. Comparison of SDM and Other Connectionist Models

SDM differs from the models proposed by Rumelhart and McClelland (1986), MacWhinney and Leinbach (1991) and Plunkett and Marchman (1991,1993) in two significant ways. First, all four connectionist models rely on error calculation. During training, the R&M PDP model for past tense verb acquisition compares its output with that supplied by the "teacher". If there is a match, no adjustments are needed, since no error exists. However, if an error is encountered, for example, the "teacher" says the output should be "1" (active) and the model computes that the output is "0" (inactive), the connection weights and thresholds are adjusted (using the perceptron convergence learning algorithm) to increase the probability that that particular unit will be active the next time the same input pattern is presented. The models implemented by M&L and P&M, since they use the back-propagation learning algorithm, also rely on the adjustment of network weights and thresholds. SDM, on the other hand, does not rely on this type (or for that matter, any type) of error calculation. This difference is significant since error calculation can be viewed as negative evidence. The language acquisition literature indicates that children do not receive a significant amount of (or any systematic) negative feedback from their parents when they incorrectly inflect a verb in the past tense (Kuczaj, 1977; Marcus et al., 1992). Since SDM does not use negative evidence (error calculation) and the other connectionist models do, SDM is more similar to the situation encountered by children as they acquire the past tense, than the connectionist models. Secondly, SDM, according to Kanerva (1988), has " a hidden layer with fixed coefficients that maps input patterns to subsets of storage locations." This allows storage and retrieval to take place only in the currently active subset of locations, and leaves most of the locations inactive at any given time. Rumelhart and McClelland's PDP model for past tense verb acquisition, on the other hand, uses the perceptron convergence learning algorithm in a network in which all layers participate in the learning process. In such a system, all units are active to varying degrees, as opposed to just being on or off, as they are in the SDM model. This difference also becomes evident when SDM is compared to other connectionist networks which use back-propagation and hidden layers (Plunkett & Marchman, 1991, 1993; MacWhinney & Leinbach, 1991). McCloskey and Cohen (1989) and Ratcliff (1990) have pointed out that the learning algorithms used in the connectionist models, result in catastrophic interference. When this type of network is first presented with a list of items to learn, it can do so with a high level of success. However, if the network is subsequently presented with a new item (or list of items) to learn, and the original items are not presented along with the new items, the success rate for the original items decreases to a very low level. This type of interference is a consequence of the fact that all units in the system are activated for each input. When new items are presented, in essence they "wipe out" the activation levels which allowed successful identification of the old items, replacing them with new activation levels based on the new items to be learned and resulting in poor subsequent performance on the original items. SDM, since it does not activate all units, but rather only activates a subset of the units, is much less subject to this type of interference.

However, even taking into consideration the above differences, it should be obvious that SDM is similar to the connectionist models in a number of ways. First, all systems rely on distributed representations of their input and mathematical reconstruction of the contents of many locations for retrieval. Secondly, although not previously mentioned, SDM also relies on parallel processing. In SDM, the comparison between input addresses and "hard" memory locations is carried out in parallel, rather than serially. Finally, and most importantly to this discussion, just as the connectionist models do not require, or use, explicit rules, SDM does not allow for rules to be explicitly encoded, it merely stores inputs and allows for their retrieval through statistical reconstruction. So, if SDM can be shown to overregularize, and to do so in a U-shaped developmental sequence, it would lend further support to the idea that explicit rules may not be required, or possibly even used, by children as they acquire language, or by anyone as they learn systems which seem to incorporate rules and exceptions.

II. Experiments and SDM Simulations

With this goal in mind, the body of this work begins with an explanation of the implementation of SDM used in all the simulations reported here. Section A will then cover the SDM simulations of Rumelhart and McClelland's (1986) "Rule of 78". Section B reexamines Palermo and Howe's (1970) experimental analogy to past tense acquisition and the SDM simulations of P&H's experiment. Finally, in section C, the details of a new experiment with human subjects, the "Friends and Enemies" (F&E) experiment, will be covered, along with the SDM simulations of this experiment.

The implementation of SDM used in this work was written by Martin Chodorow and will hereafter be referred to as "the program". Any simulation run using the program initially requires values for the following parameters: 1) the length of address and data vectors, 2) the number of "hard" locations, and 3) the seed. Once selected, these parameters are fixed for the duration of the simulation and are used to determine the memory space. The vector length specifies the number of bits in the address vector which is used to store and retrieve all the information in the simulation. This value also determines the total memory space. For example, if the vector length is set to 16, then the memory space is 2^{16} = 65,536 addresses. Kanerva (1988) suggests that vectors of length 1000 might be necessary to approach the richness of human memory representation. Fortunately, the essential properties of SDM can be observed with much shorter vectors (Denning, 1989; Keeler, 1988), and in the simulations reported here, vector lengths between 16 and 64 are used. An important feature of Kanerva's (1988) theory stipulates that only a small number of addresses (relative to the total memory space) need to be available for memory storage and retrieval (hence the "sparse" in Sparse Distributed Memory). In the following simulations, the number of these "hard" locations varies between 2000 and 5,000. Finally, the program requires a seed. The seed is used by the computer's random number generator to produce the "hard" location addresses. It is also used for the production of noise, as described below. If no seed (or a

"0") is entered, the program will provide a seed using the computer's timer chip.

During each simulation, a number of additional parameters must also be specified, however, unlike the parameters indicated above, these parameters may vary during the course of the simulation. They include: 1) the address vector which is used during storage or retrieval from memory, 2) the data vector, used to represent the information stored, 3) the radius, which determines the neighborhood of hard locations from which (or to which) data will be read (or stored), and 4) noise, used to distort address vectors. The address and data vectors are unique to each simulation and will be described in more detail in the following sections. Kanerva (1988) recommends that the radius, (i.e., Hamming distance) used to store and retrieve information in SDM should encompass 1/1000th of the total memory space. Based on the normal distribution approximation to the binomial, this is equivalent to a value about three standard deviations away from the mean of the Hamming distances between addresses in the memory space. In all the simulations run for this dissertation, the radius was calculated using the following formula:

$$\text{radius} = (N/2) - 3(\sqrt{N}/2)$$

where N = the vector size, $N/2$ = the mean Hamming distance from any given address to any other address in the memory space, and $\sqrt{N}/2$ = the standard deviation of the distribution of the Hamming distances. Since human memory storage and retrieval is rarely performed under ideal conditions in which there are no distractions or interference, noise is frequently used in these simulations to store and retrieve from distorted address vectors. The program adds noise probabilistically. The noise value may be between zero and one and indicates the probability that a bit in the vector will be randomly replaced with its complement (i.e., "1" would be replaced with "0" or "0" with "1"). Noise values between 0 and 0.30 were used in the simulations performed as part of this work.

Since the effects of manipulating the radius and noise values may not be readily apparent, further explanation seems appropriate. Although the memory space in SDM is multidimensional, a two dimensional space will suffice to illustrate the effects. Suppose two items (data vectors), a and b, are to be stored in SDM at two different addresses, **A** and B. The radii, r_a and r_b used to store the items, can be thought of as specifying circular storage areas in memory surrounding addresses **A** and B, respectively. Since SDM stores in a distributed manner, the data vectors representing a and b will be stored at all hard locations (addresses) in the memory space around **A** and B, within radii r_a and r_b respectively. If the distance (d) between A and B is larger than the sum of r_a and r_b there will be no overlap between the two storage areas, and SDM will have no trouble distinguishing between a and b upon retrieval (see Figure 4a). If, however, d is kept constant, but r_a and r_b are increased sufficiently, the storage areas around A and B will overlap. When SDM tries to retrieve either a or b in this situation, it may have more trouble distinguishing between the two since they share part of their storage space (see Figure 4b). The addition of noise to the specified address vectors, has a similar effect. When noise is added, the address changes slightly and results in a storage area slightly off center from A or B. Since noise is added randomly, a slightly different storage area in the memory space is used each time. Over a number of trials, the addition of noise, in effect, increases the storage area around the original address vector. If d, r_a, and r_b are kept constant, and enough noise is added, the storage areas around A and B will overlap more, and SDM may have difficulty distinguishing between a and b upon retrieval (see Figure 5).

A. "Rule of 78" Simulations

1. Method

First, in order to demonstrate that SDM can overregularize and do so in a U-shaped developmental sequence, Rumelhart and McClelland's (1986) "rule of 78" was simulated using the program.

The seeds used to initialize the memory space were randomly generated by the program in each simulation. In

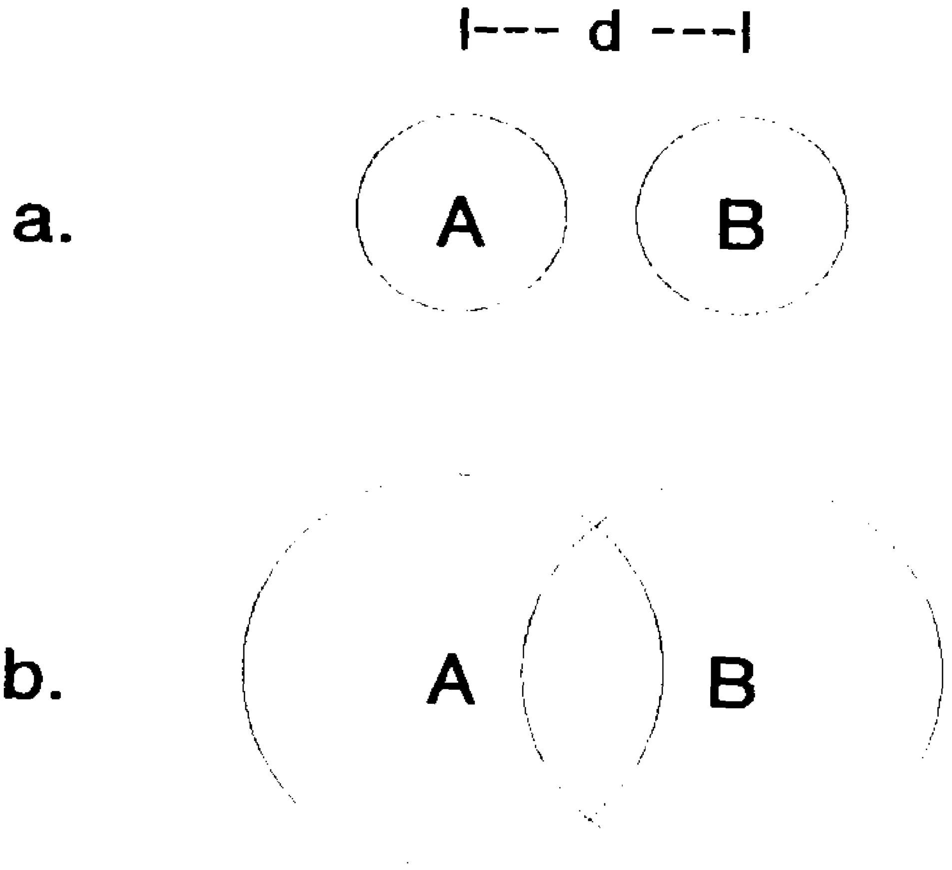


Figure 4. Effect of increasing the radii on the storage areas around addresses **A** and **B**. **d** = Hamming distance between **A** and **B**.

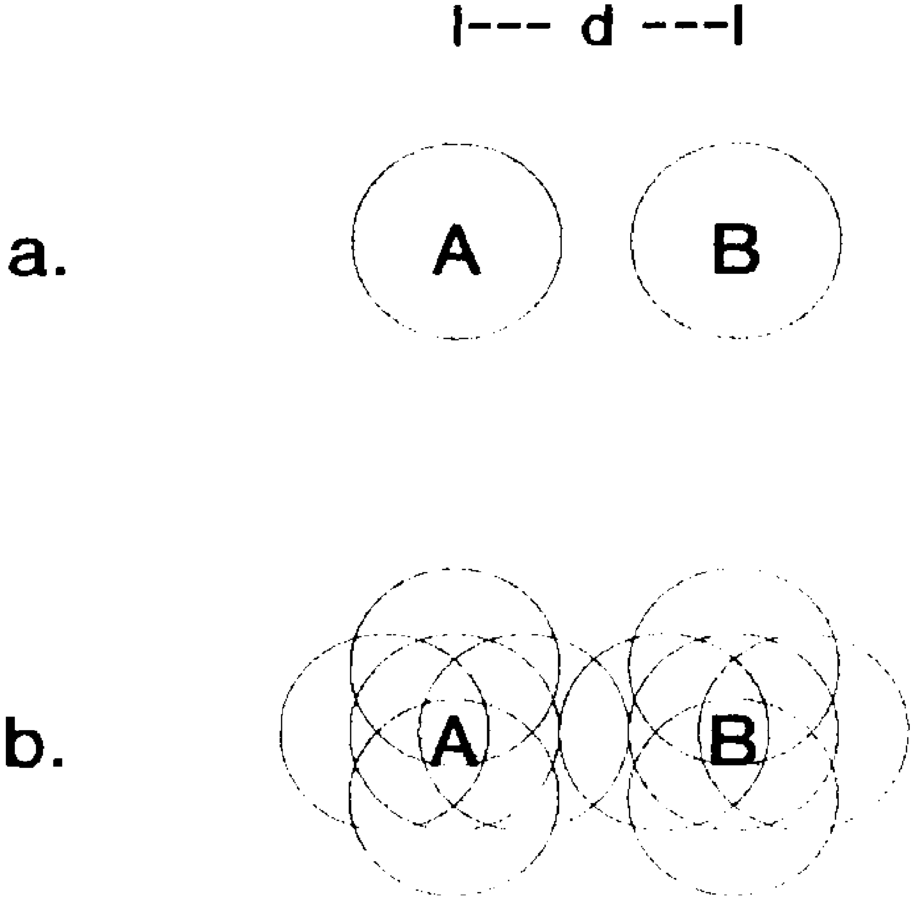


Figure 5. Effect of increasing noise on the storage areas around addresses **A** and B. d = Hamming distance between **A** and B.

all the simulations of the "rule of 78", the vector representing the response was stored at the vector address of the stimulus. In this way, during testing, if the association between the stimulus and the response is learned, the stimulus address should contain the bit vector representing the response. Sixteen-bit vectors were used to represent both the stimuli and the responses (see Figure 6). For example, the response "248", which was represented by the vector "0101000101010001", was stored at the vector address "0101001001010010", which represents its stimulus, "247". An example of the graphical representation of these vectors, used by the program is shown in Figure 7. Prior to the addition of noise, all digits were coded with non-overlapping bits in the address and data vectors. Two of the 16 bits were used for each of the digits 1-8. For example, the bit positions used to represent "4" were different from those used for "5". Five thousand "hard" memory locations were used. During training the responses were stored at all addresses within a radius of 2 from the designated stimulus address. Similarly, responses were read out of all addresses within a radius of 2 from the designated stimulus address during testing. A radius of 2 was chosen to include 1/1000th of the total memory space, using the formula in the previous section.

Ten simulations, which differed only in the seed used to initialize the memory space, were performed. Each

REGULAR PAIRS

Stimulus	Response
148	147
157	158
158	157
167	168
168	167
247	248
248	247
257	258
258	257
267	268
268	267
347	348
348	347
357	358
358	357
367	368
368	367

IRREGULAR PAIR

Stimulus	Response
147	147

OVERREGULARIZATION

Stimulus	Response
147	148

Figure 6. Rule of 78 stimulus - response pairs. The overregularization pair was not presented during training, it is listed only for the purpose of clarification.

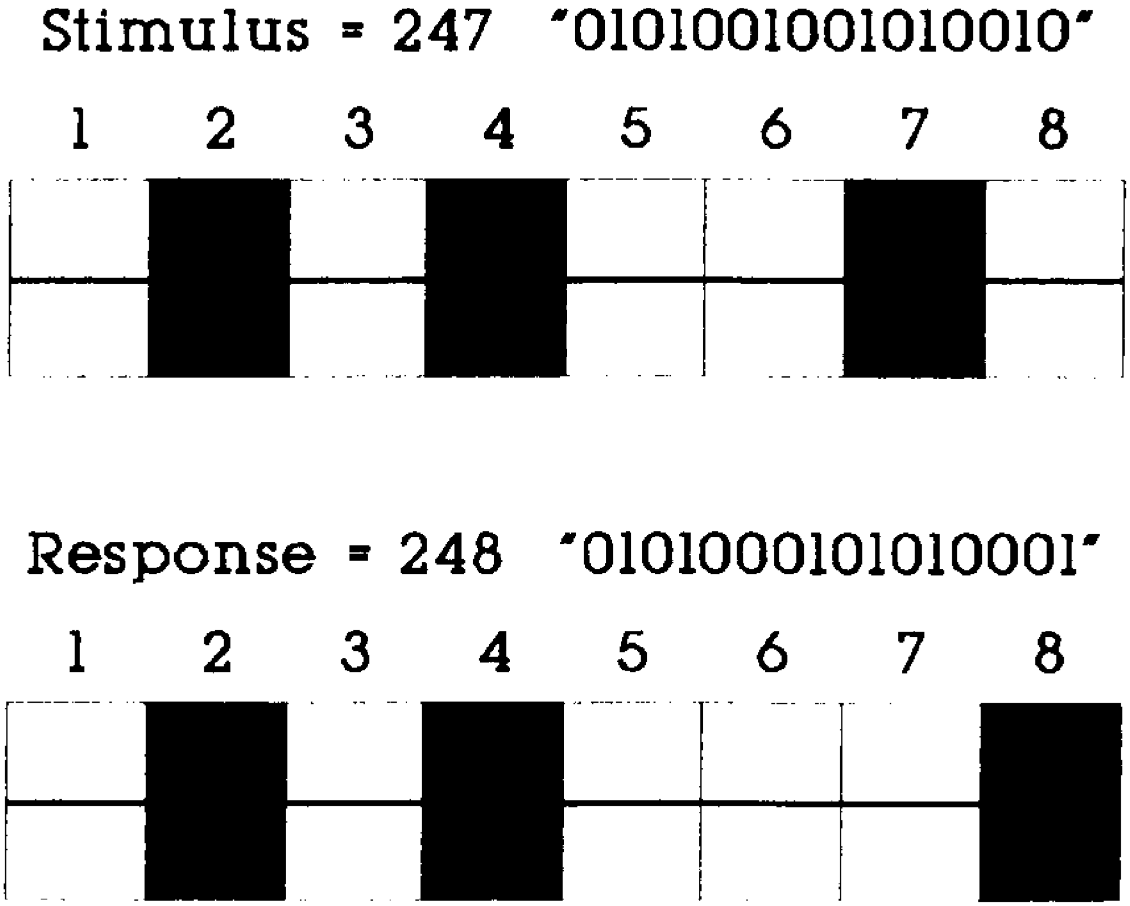


Figure 7. Rule of 78: Graphical representation of the stimulus-response vectors used by the program in the SDM simulations of the rule of 78. Two rows of 8 boxes are used to represent the 16 bit vectors. Each column of boxes represents the two bits used to represent one of the 8 possible digits in the stimulus or response (labelled for the purpose of illustration). A black box represents a "1" in the vector and indicates that the bit is turned on. A white box represents a "0" in the vector and indicates that the bit is off.

simulation consisted of 50 trials. Each trial consisted of a training session followed immediately by a testing session. During training, each of the 17 regular pairs was presented one time and the irregular pair was presented two times. During each testing session, the data contained in each of the 18 stimulus addresses were read five times. During both training and testing sessions the noise level added to each stimulus address vector was initially set to 0.20. This means that for each bit there was a probability of 0.20 that it would be replaced with its complement. After every two trials this noise level was reduced by 0.01 until, during training, the level reached 0.02, and during testing, the noise level reached 0.0. Therefore, during training, trials 37 through 50 had a noise level of 0.02 and during testing, trials 41 through 50 had no noise added to the stimulus address. See Table 1 for a summary of the parameters used in these simulations.

2. Results

The results of the rule of 78 simulations, averaged by trial over the ten simulations, are presented in Figure 8. Throughout the course of the simulations, the irregular response was identified correctly more often than the seventeen regular responses. The irregular reached perfect performance for the first time at trial 24 and was consistently perfect by trial 34, while the regulars did not reach their peak performance (94.1% correct averaged over

Table 1

Parameters for the "Rule of 78" Simulations

Vector length	16 bits
Number of hard memory locations	5000
Seed	randomly generated by the program
Number of trials per simulation	50
Training:	
Stimulus-Response presentations	17 regular pairs each presented 1x/trial 1 irregular pair presented 2x/trial
Radius	2
Noise added to stimulus address vectors	Trial 1: Noise = .20, and is reduced by .01 after every 2 trials until Noise = .02 (trials 37-50)
Testing:	
Response readings (from stimulus address)	17 regular responses each read 5x/trial 1 irregular response read 5x/trial
Radius	2
Noise added to stimulus address vectors	Trial 1: Noise = .20, and is reduced by .01 after every 2 trials until Noise = .00 (trials 41-50)

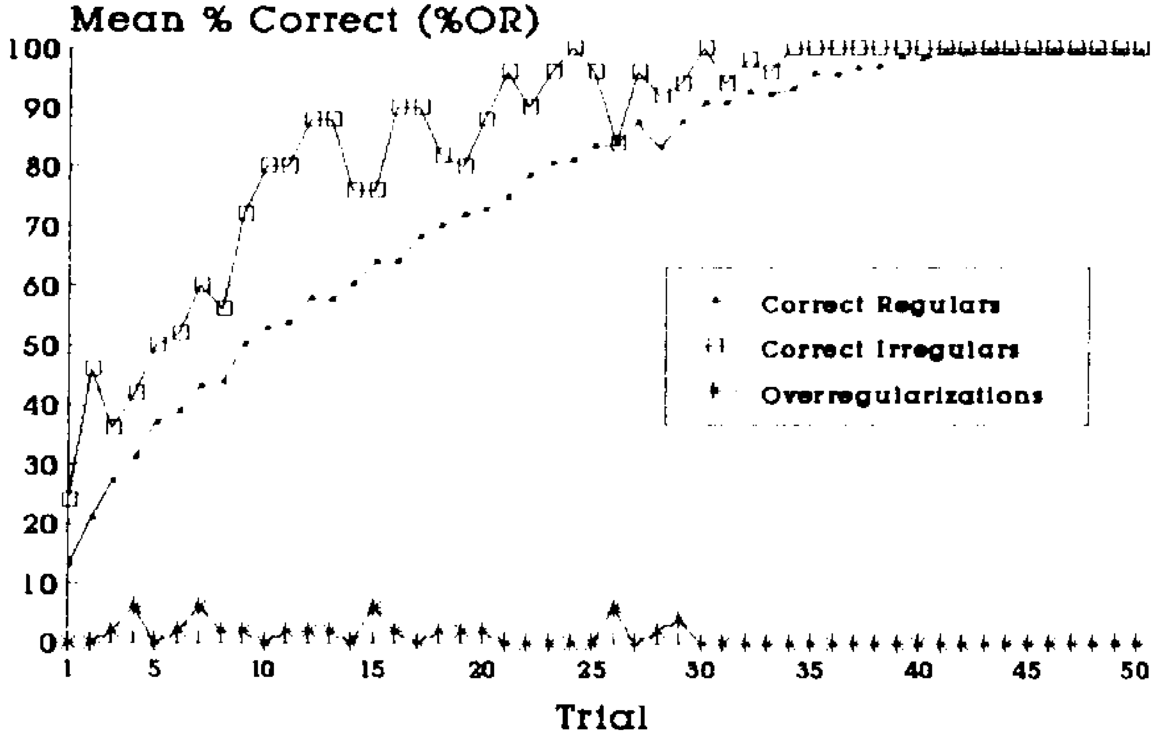


Figure 8. Mean results by trial for the 10 SDM simulations of the rule of 78. The mean percent correct is presented for the regular and irregular pairs. Overregularization errors (%OR) apply only to the irregular.

all regular pairs in all 10 simulations) until trial 41. Low levels of overregularization (between 2% and 6%) can be seen between trials 4 and 29.

Since the time course of acquisition for the irregular pair in each of the ten simulations is different, it not clear whether the simulations showed a U-shaped developmental sequence. Additionally, none of the simulations started out performing perfectly on the irregular. On average, it took ten trials for the simulations to reach 100% correct on the irregular, with the individual simulations varying between five and sixteen trials. In order to equate the simulations, Figure 9 shows the average results for the ten simulations beginning at the trial in which each simulation first reached 100% correct on the irregular. Here the U-shaped developmental sequence becomes evident. Averaging the data and presenting it in blocks of five trials makes the U-shape even more pronounced (see Figure 10)[5]. If one considers the point at which the irregular first reaches 100% correct as the beginning of the developmental sequence, then the irregular clearly shows a period of declining performance followed by a subsequent

---

[5] Appendix A includes Figures AP-A1 through AP-A10 which show each of the ten simulations graphed in the same way as Figure 9.

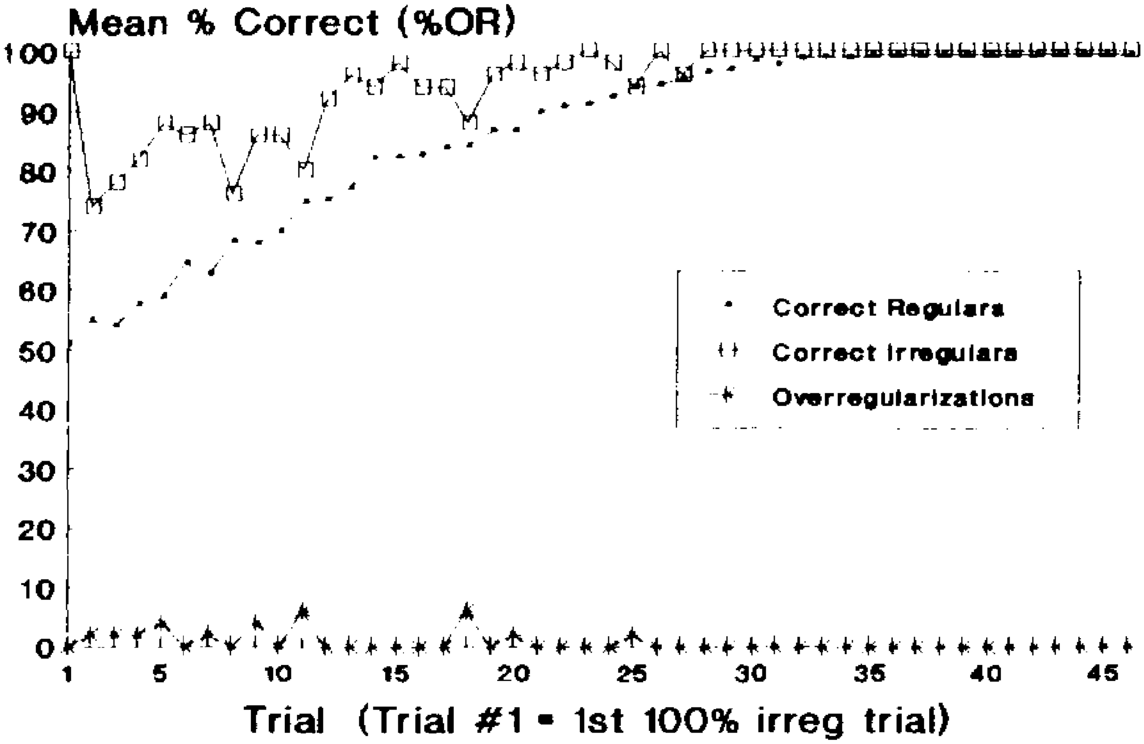


Figure 9. Mean results by trial for the 10 SDM simulations of the rule of 78, starting at the trial in which each simulation first reached 100% correct on the irregular. Overregularization errors (%OR) apply only to the irregular.

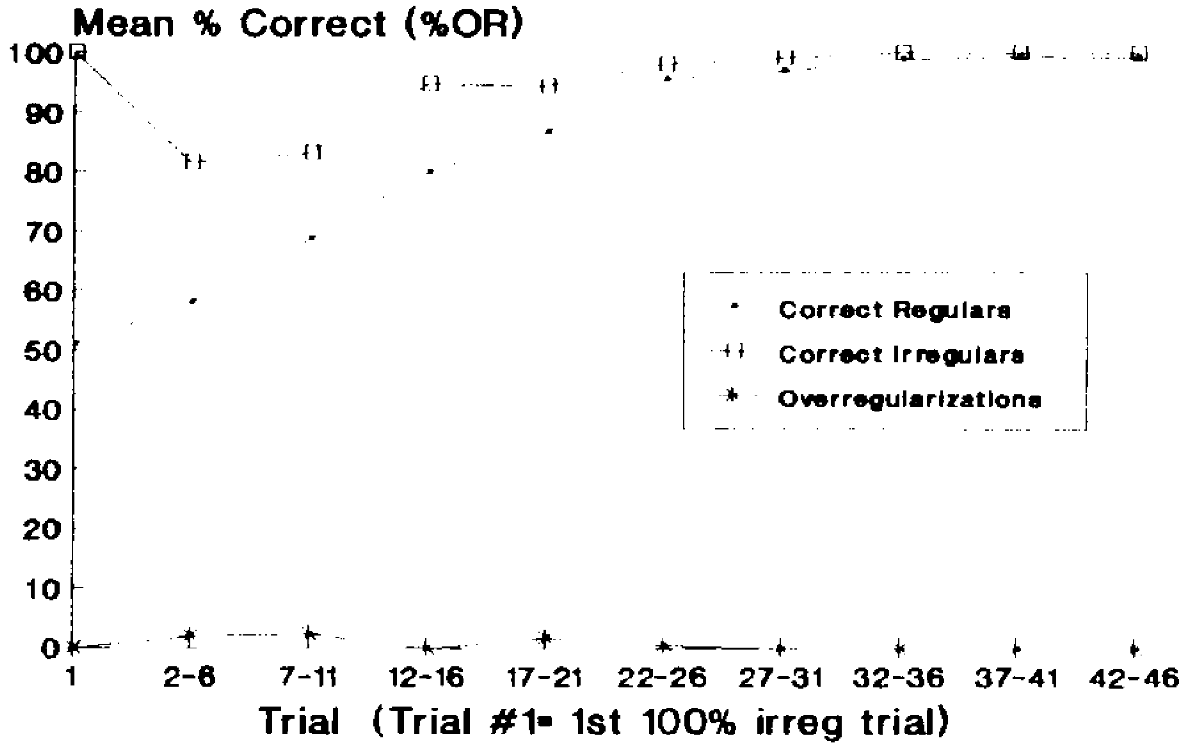


Figure 10. Mean results by blocks of 5 trials for the 10 SDM simulations of the rule of 78, starting at the trial in which each simulation first reached 100% correct on the irregular. Overregularization errors (%OR) apply only to the irregular.

increase in performance.[6] During this period, low levels of overregularization are present. The regulars, on the other hand, do not reach perfect or near perfect performance until late in the simulations and do not show any U-shaped learning. Instead, the regulars exhibit steady improvement over time. In essence, when presented with a set of stimulus-response pairs which include pairs which follow a rule and one which does not, SDM exhibits both overregularization and U-shaped learning. It does so without making any distinction between regular and irregular pairs, and without explicitly encoding the rule of 78.

3. Discussion

Rumelhart and McClelland (1986) used the rule of 78 to

---

[6] Some may question the decision to consider the point at which the SDM simulations first reach 100% on the irregulars, rather than the first trial of the simulations, as the point which is analogous to the beginning of the past tense acquisition curves for children, since it has been claimed that when children begin inflecting the past tense, they do so correctly, if at all. This decision was based on a number of factors including: 1) R&M's simulations of the rule of 78 only reached near perfect performance on the irregulars after 20 trials, 2) Plunkett & Marchman's (1988) past tense acquisition network also did not start out performing perfectly on their three classes of irregulars, 3)R&M's and Plunkett & Marchman's simulations only showed U-shaped learning after the irregulars first reached perfect or near perfect performance and, most importantly, 4) it is unclear that the language acquisition data obtained from children actually picks up children's first attempts at past tense inflection. It seems highly unlikely that children begin using the past tense for irregulars perfectly. Instead, it seems more likely that a child's first attempts are not correct, but cannot easily be recognized as attempts to use the past tense and are therefore not identified as errors.

demonstrate that their connectionist network could learn, in a way similar to children learning the past tense of verbs, a simple set of stimulus-response pairs which include both regular and irregular pairs. The goal of the SDM simulations detailed above was the same. As indicated previously (see section I.D.1.a.), the R&M model showed U-shaped development and overregularization in the process of learning the rule of 78. However, the SDM model of the rule of 78 has several distinct advantages over the R&M model. First, the R&M model used a two phase input process which biased its output. The initial input to the R&M model consisted of one regular pair (258->257) and the irregular pair (147->147). These two pairs were presented for 20 trials. By the end of these trials, the model's performance was nearly 90% correct on both pairs. According to Rumelhart and McClelland, this initial training phase represents the situation a child encounters early in the past tense acquisition process. In this early stage a child only encounters a few high frequency words, and since the majority of the most frequent verbs in English are irregular (Kucera & Francis, 1967), this initial phase is dominated by irregulars. In the second phase of their training, R&M presented all 18 pairs (17 regular and the irregular) for an additional 190 trials. According to R&M, as children proceed through the acquisition process and more verbs are learned, the proportion of regular verbs in their vocabulary increases. This phase of a child's acquisition process corresponds to the introduction of the other 17 regular pairs to the R&M model.

Pinker and Prince (1988) point out that children do not encounter verbs in any way comparable to R&M's discontinuous input set (R&M's input can be considered discontinuous since initially only two pairs are presented to the network and only after a high level of performance is reached on these two pairs, one of which is the only irregular in the input set, is the network exposed to all pairs). Although directed at Rumelhart and McClelland's more elaborate network which uses phonological representations of verbs, their criticism is equally relevant here, since R&M present the rule of 78 as an analogy to past tense acquisition and present similarly discontinuous input sets in both network models. Pinker and Prince present data from four children which clearly indicate that children do not necessarily evidence an explosive vocabulary growth at any point in past tense acquisition (see Pinker & Prince, 1988 - Figure 2, p.141). Not only is R&M's input set different from that experienced by children, but, P&P argue, it is this discontinuous input set which causes the R&M model to overregularize and show a U-shaped learning curve for the irregulars. They therefore question the suitability of the R&M model as a representation of past tense acquisition. Lachter and Bever (1988), in their similarly extensive review of the R&M model, voice this same criticism. Like Pinker and Prince, they could find no evidence of a sudden increase in the number of verbs children know when they begin to overregularize.

The SDM simulations performed as part of this work do not use a discontinuous input set and as a result, have a clear advantage over the R&M model. Since the input training set remains constant throughout the simulations, the decrease in performance on irregulars and the corresponding appearance of overregularization errors in the SDM simulations cannot be attributed to a change in the input set. Of course the argument could be made that the input set used in these simulations is subtly changed since noise is added in decreasing increments throughout training. However, as will be more fully discussed in the ensuing general discussion, there are several possible psychological equivalents of the noise used in these SDM simulations. These include extrinsic noise encountered by children as they acquire language, intrinsic noise which places a cognitive load on the child, and a "noisy" nervous system which is busy myelinating its fibers throughout childhood. Although drastically simplified, the training pairs used in the SDM simulations are also more similar, in one critical way, to those encountered by children in past tense acquisition than those used by R&M in their demonstration of the rule of 78. As previously noted, most of the most frequent verbs in English are irregular (Kucera & Francis, 1967). Instead of representing this by only training the model exclusively with one regular and one irregular initially, and then presenting the rest of the pairs, as R&M did, the SDM simulations performed here present the irregular twice as frequently as each regular pair throughout training.[7] In this way the frequency element encountered by children is addressed directly.

A second advantage these SDM simulations have over the R&M model is due to the intrinsic differences between the models. As previously noted (section I.D.5), the R&M network relies on a "teacher" and error calculation and SDM does not. In the R&M model, output is compared with the "teacher's" correct output, and if there is a discrepancy, connection weights and thresholds are adjusted to increase the probability of a correct output next time. SDM does not require anything similar to a "teacher". Children do not appear to learn the past tense in any formal way. The developmental period in question usually occurs before the age of five when children traditionally begin their formal

---

[7] The decision to present the irregular twice per trial was an empirical one. Keeping all other parameters constant, simulations were performed in which the irregular was presented three, four and five times per trial. When the irregular was presented more frequently than twice per trial, irregular performance was almost immediately perfect, with no evidence of U-shaped learning or OR errors. This frequency effect is consistent with results obtained by Palermo and Howe (1970) and will be addressed more fully in the general discussion.

education. Therefore, SDM, which doesn't require explicit teaching, is more analogous to the informal learning process employed by children when they learn the past tense. In addition, the error calculation and correction used in the R&M model may be viewed as a kind of negative feedback. That is, the "teacher" tells the model it is wrong and the model makes the appropriate corrective measures. The past tense acquisition literature indicates that children do not receive any significant amount of negative feedback from their parents (Kuczaj, 1977; Marcus et al., 1992). When children incorrectly inflect a verb in the past tense, parents do not usually correct them, and if they do, the correction is not done in any systematic way. Since SDM does not use any type of error correction, it more closely resembles the situation encountered by children than the R&M model does.

It would seem therefore, that SDM is superior to the R&M rule of 78 network in its ability to exemplify the past tense acquisition process. However, the question remains as to how well the SDM model compares to the actual child data. Marcus and his colleagues (1992) use the data provided by many language acquisition researchers (made available by the Child Language Data Exchange System) and they reanalyze the data in a variety of ways related to U-shaped learning and overregularization. If the results of the simulations performed using SDM compare well with the analysis provided by Marcus et al., an even stronger case may be made for SDM's ability to model past tense acquisition.

There has been little consensus in the language acquisition literature about how often children overregularize. Studies of individual children cite rates as high as 45% overregularization (Kuczaj, 1977). Some have even claimed that when children begin to overregularize, they do so exclusively and therefore, for a period, the overregularization rate is 100% (Bowerman, 1982 as cited in Marcus et al., 1992). The work by Marcus and his colleagues attempts to quantify just how often children do overregularize. The first thing they looked at was the overall overregularization (OR) rate which they have defined as:

$$\begin{array}{l}\text{Overall}\\ \text{OR Rate}\end{array} = \frac{\text{\# OR tokens}}{\text{\# OR tokens} + \text{\# correct irregular past tokens}}$$

The median overall OR rate for 25 children with individual transcripts was found to be 2.5% and the mean was 4.2%. The individual children's OR rates ranged from 0% to 24%. Most of the children's OR rates fell well below 10%, with only two exceptions, Abe at 24% and April at 13% (see Marcus et al., 1992, Table 2). The SDM simulations performed as part of this work also show low levels of overregularizations (see Table 2). The median OR rate for the ten simulations performed was 1.40%, the mean was 1.20% and the range was 0% to 2.35%. Although these values are lower than those

Table 2

Overregularization (OR) Rates for Individual Simulations of the "rule of 78"

Simulation	# Correct Irregulars	# OR	OR Rate(%)
DX	210	3	1.41
PX	214	2	0.93
QX	219	0	0.00
RX	208	5	2.35
TX	207	3	1.43
UX	205	3	1.44
VX	226	3	1.31
WX	215	3	1.38
XX	220	0	0.00
YX	220	4	1.79
Mean of the 10 Individual Simulations			1.20
ZX[8]	240	9	3.61

---

[8] Pilot simulation ZX used the same parameters as simulation RX except the initial noise level was set at .30. In addition, because of the higher noise level, 70 trials were required to reach peak performance, rather than the 50 trials required in simulation RX.

reported in the language acquisition data, they are consistent with the claim that OR rates are very low. In addition, OR rates for children as indicated above include overregularizations for many irregular verbs, while the SDM simulations only include one irregular. The difference between OR rates in children and the SDM simulations can, at least partly, be attributed to the difference in the richness of the "vocabularies" each is exposed to. One way to compensate for this difference is to increase the amount of initial noise added to the address vectors in the simulations. More initial noise should result in higher levels of overregularization. To address this possibility, a pilot simulation (ZX), which used the same parameters as one of the simulations detailed above (RX), was performed using a starting noise level of 0.30. Table 2 indicates that when compared to the original simulation RX, which employed an initial noise level of 0.20, pilot simulation ZX showed an OR rate 52% higher than simulation RX. Therefore, had initial noise levels been set higher in these SDM simulations, the mean OR rate would probably have been closer to that reported in the child acquisition literature.

Another issue addressed by Marcus et al., (1992) is the fact that different verbs have been shown to have different OR rates over time. Marcus and his associates present figures representing four types of verbs classified by their pattern of overregularization (Marcus et al., 1992, figures

12-15). One classification consists of verbs which are rarely overregularized and includes the verbs say, find, forget, see and tell. Figure 11 compares the overregularization patterns of these verbs (Figure 11a) to that of one of the SDM simulations of the rule of 78 (Figure 11b)[9]. Clearly the patterns are very similar.

As a final comparison between the child acquisition data and these SDM simulations, the data for one child, Adam from the Brown (1973) study, were compared to the mean results for the ten SDM rule of 78 simulations. Figure 12a shows Adam's overregularization rate (presented as 100 - OR rate) over time. The comparison to the SDM simulations (Figure 12b) is striking. Although Adam showed a longer initial period of perfect irregular performance and does not demonstrate the portion of the right hand arm of the U in which perfect performance is again achieved, the period during which both Adam and the simulations overregularized is very similar.

In summary, by simulating Rumelhart and McClelland's (1986) rule of 78, it has been demonstrated that SDM can overregularize and show U-shaped learning similar to that shown by children as they acquire the past tense. In doing so, SDM was shown to more closely parallel claims made about how children acquire the past tense than the R&M rule of 78

---

[9] The data for the verb classification including say, find, forget, see and tell is taken from one child, Abe, in Kuczaj's (1977) study.

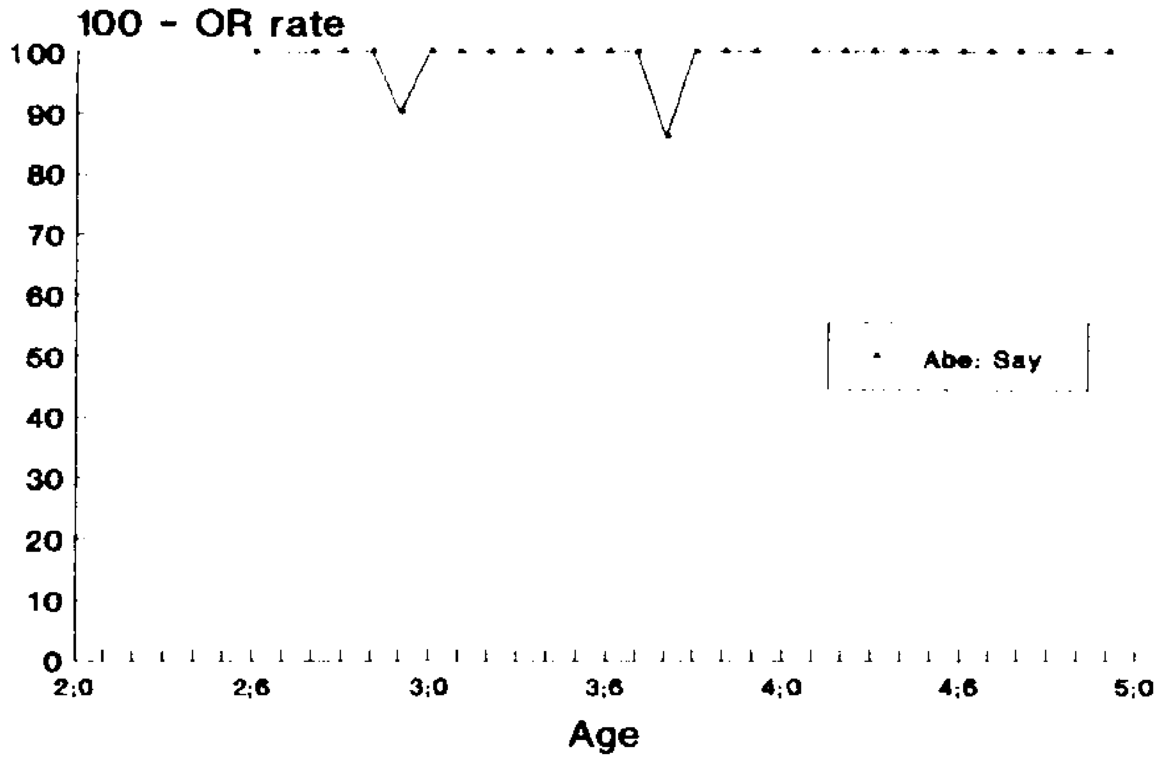


Figure 11a. Abe's pattern of overregularization for a class of irregular verbs which is rarely overregularized. This verb class includes the irregular verbs, _say_, _find_, _forget_, _see_ and _tell_.

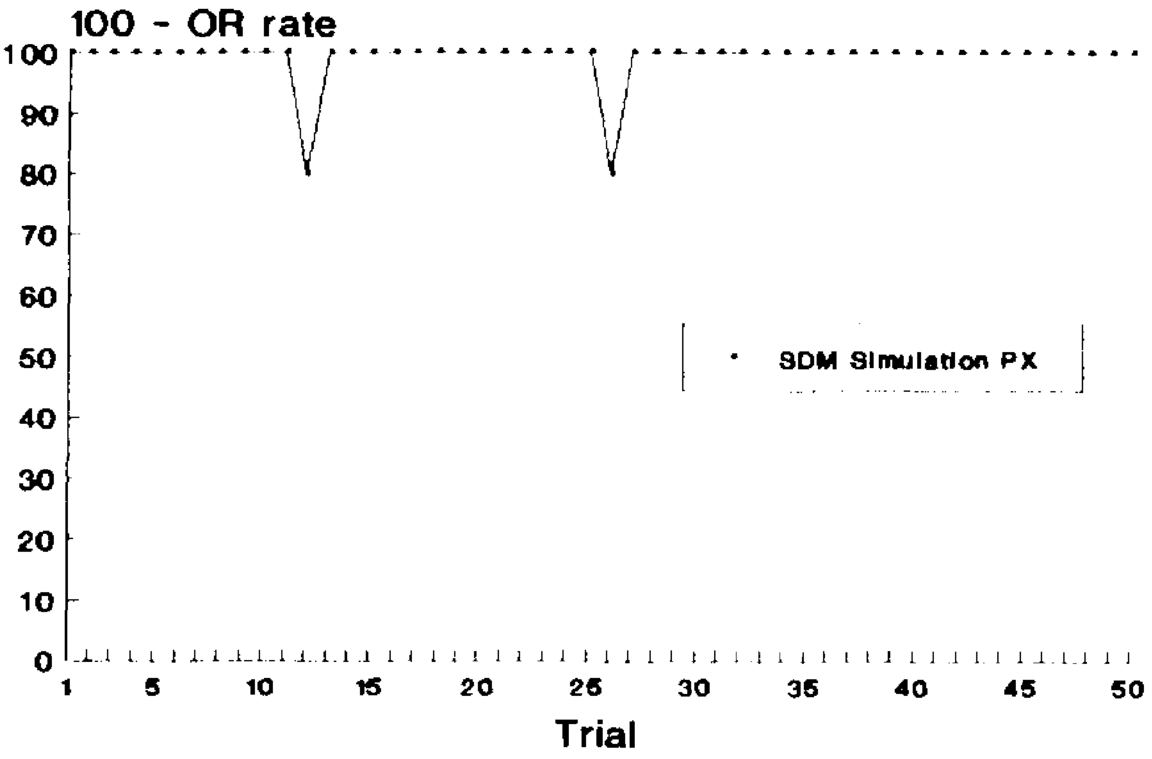


Figure 11b. The pattern of overregularization for one SDM simulation of the rule of 78 (simulation PX).

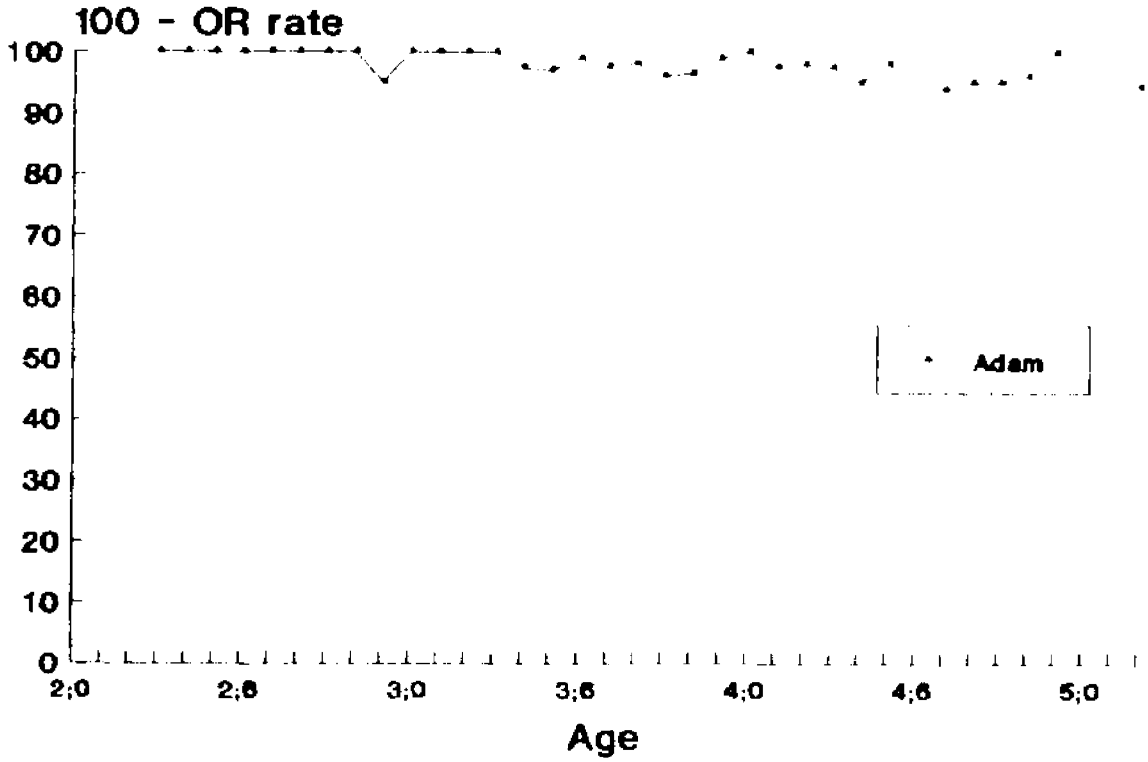


Figure 12a. Adam's overregularization rate over time for all irregular verbs.

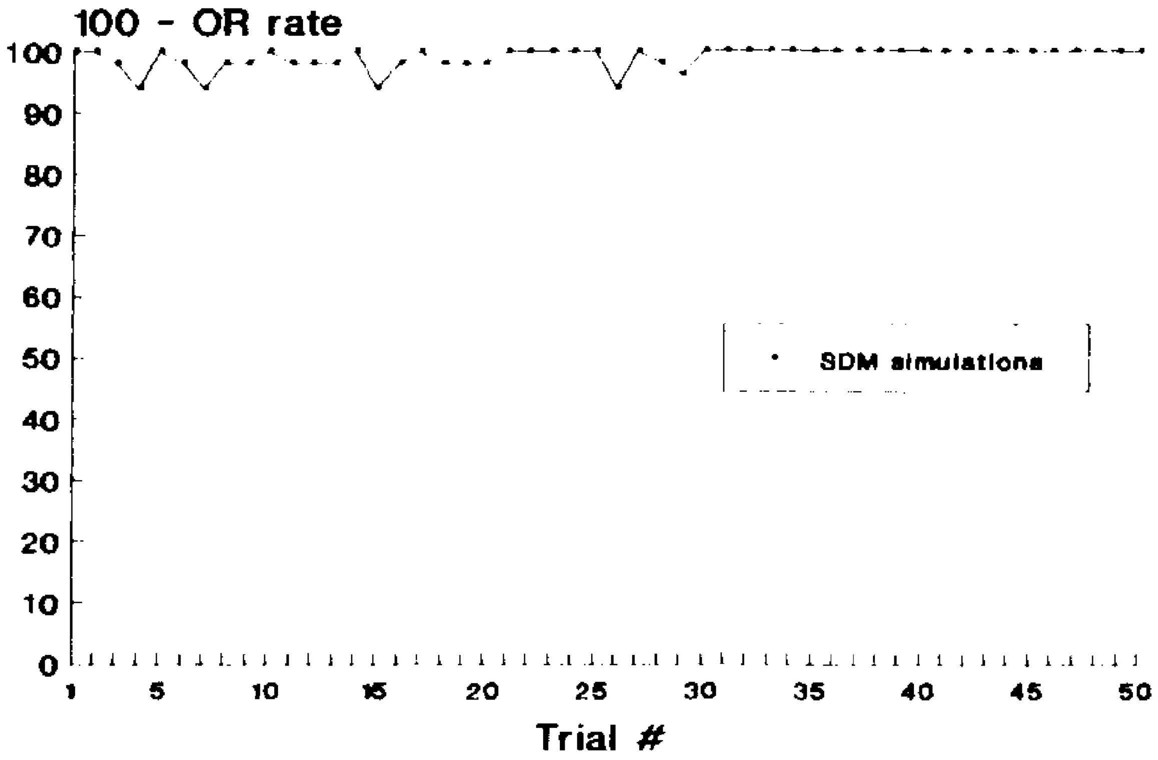


Figure 12b. The mean overregularization rate over time for the 10 SDM simulations of the rule of 78.

network in three significant ways. First, the SDM simulations reported here present input which is more similar to the input a child receives than the R&M model does in that the SDM simulations directly address the fact that irregular verbs are more frequent than regular verbs. Second, the input to the SDM model was not discontinuous, as was the R&M input. Third, the R&M model requires a "teacher" which provides error correction to increase its performance, while SDM and children do not. Children do not appear to need this type of negative feedback, and SDM does not utilize it. Finally, SDM is not only superior to the R&M model in its handling of the overregularization and U-shaped learning phenomena, it also shows results which closely resemble the data in the child acquisition literature.

B. Palermo and Howe (1970) & Simulations

As a further attempt to establish the viability of SDM as a model of human memory, Palermo and Howe's (1970) (hereafter referred to as P&H) experiment was simulated using the program. Before discussing the simulations, P&H's experiment is summarized and their data reanalyzed. The goals of this reanalysis include: 1) ascertaining the difference in OR rates among the irregulars, based on the frequency of presentation of each irregular pair, and 2) revealing any U-shaped learning among P&H's subjects.[10]

1. Palermo and Howe (1970)

a. Method

As previously discussed (see section I.C.1 & 2), in an early study, Palermo and Eberhart (1968) used a rule learning task to demonstrate that the behavior of children acquiring the past tense could be simulated in a laboratory setting. They found that irregular instances of the rule were learned before regular cases and that later in the learning process, there were overregularizations of previously learned irregulars. In 1970, P&H attempted to improve the analogy between past tense acquisition and their experiment by changing the modality of stimuli and responses and by making their rule based and irregular pairs more sensitive to the linguistic characteristics of past tense acquisition.

P&H used four groups of six subjects each. The groups were presented with different "regular" and "irregular" pairs. For all groups, the regular stimuli were randomly selected two-digit numbers, with the constraint that on each trial at least one number end in each of the digits 1-9. Groups A1 and B1 required responses of "F", "G" and "C" for regular stimuli and Groups A2 and B2 required responses of "O", "I" and "A" for their regular stimuli. The second

---

[10] I am grateful to David Palermo for kindly providing the raw data necessary for this reanalysis.

digit of the two digit stimulus determined the response for each pair. The difference between the consonant response groups (A1 and B1) was in the responses required to each specific second digit. For example, Group A1 required a response of "G" to all two-digit stimuli ending in either "3", "4" or "8" while Group B1 required the same response ("G") for all stimuli ending in either "1", "4" or "6". This difference was also seen in the vowel response groups (A2 and B2). Irregular pairs required vowel responses ("A", "I", "O" or "U") in the regular consonant response groups (A1 and B1) and similarly, irregular pairs required consonant responses ("P", "Q", "R" or "T") in the regular vowel response groups (A2 and B2). The difference in irregulars between the groups (A1 and B1; A2 and B2) was both in the particular response required for each two-digit stimulus and in the number of presentations per trial of each particular irregular pair (in the study trial) or stimuli (during the anticipation trials). For example, Group A2 was presented with the irregular pair "84-T" four times, "46-P" three times, "57-Q" twice and "71-R" once per study trial, while Group B2 was presented with "93-Q" four times, "42-T" three times, "85-P" twice, and "26-R" once per study trial (see Table 3).

The experiment began with a study trial during which all 22 pairs (12 regular, 10 irregular) were presented to each subject. During this trial, the subjects were

Table 3

Characteristics of the Stimulus - Response Sets used in Palmero and Howe (1970)

	Regular Pairs		Irregular Pairs	
Group	Last Digit of Stimulus	Response	Stimulus	Response
A1	1, 9, 7	F	4X: 84	U
	3, 4, 8	G	3X: 46	I
	5, 2, 6	C	2X: 57	O
			1X: 71	A
B1	7, 5, 9	F	4X: 93	O
	1, 4, 6	G	3X: 42	A
	3, 2, 8	C	2X: 85	U
			1X: 26	I
A2	1, 9, 7	O	4X: 84	T
	3, 4, 8	I	3X: 46	P
	5, 2, 6	A	2X: 57	Q
			1X: 71	R
B2	7, 5, 9	O	4X: 93	Q
	1, 4, 6	I	3X: 42	T
	3, 2, 8	A	2X: 85	P
			1X: 26	R

presented with both the stimuli and their associated responses and were not required to respond, but instead, were instructed just to study the pairs. This study trial was followed by anticipation trials during which the subjects were presented with the 22 two-digit stimuli and were required to respond with the appropriate letter responses. All subjects received at least 25 anticipation trials and continued until a criterion of three successive errorless trials was reached or for a maximum of 35 trials.

b. Results and Reanalysis

The mean number of trials to a criterion of one errorless trial for the entire list, as well as for the regular pairs, was 26.63[11]; while the mean number of trials to criterion for the irregular cases was 9.88. Therefore, just as in Palermo and Eberhart (1968), the irregular pairs were learned well before the regular pairs. No subject reached criterion on the regulars before the irregulars. In addition, performance on the regular forms was consistently below that of all irregulars, while there was an orderly increase in acquisition rates for the irregulars based on the number of presentations per trial for each. That is,

---

[11] There is a slight discrepancy in the mean number of trials to criterion for the whole list and the regular pairs between P&H's published results and the results of the reanalysis conducted as part of this work. P&H report that the mean number of trials to criterion was 26.87 for the whole list and 26.58 for the regular pairs, while the reanalysis performed here shows both to be 26.63. This discrepancy is probably the result of only one or two data points and is therefore not considered critical.

the irregular presented four times per trial was learned more quickly than the irregular presented three times, followed by the irregular presented twice and finally the irregular presented once per trial (see Figure 13). P&H also indicated that their subjects responded to regulars at chance levels until criterion was reached on the irregulars, at which time regular response levels rose above chance (this was taken by P&H as evidence of the traditional two process theory of past tense acquisition).

P&H's subjects made a total of 256 overregularization (OR) errors throughout the experiment. The majority of these OR errors (157 or 61% of all OR errors) occurred before criterion was reached on the irregulars and 99 (39% of all OR errors) occurred after criterion was reached on the irregular pairs. OR errors accounted for 25% of all errors made on irregulars over all trials, 19% of all irregular errors before criterion was reached on the irregulars, 54% of all irregular errors after criterion was reached on the irregulars, and 100% of all irregular errors made after a criterion of one errorless trial was reached on the regular pairs. Therefore, as P&H's subjects became more proficient at correctly responding to regular stimuli, the

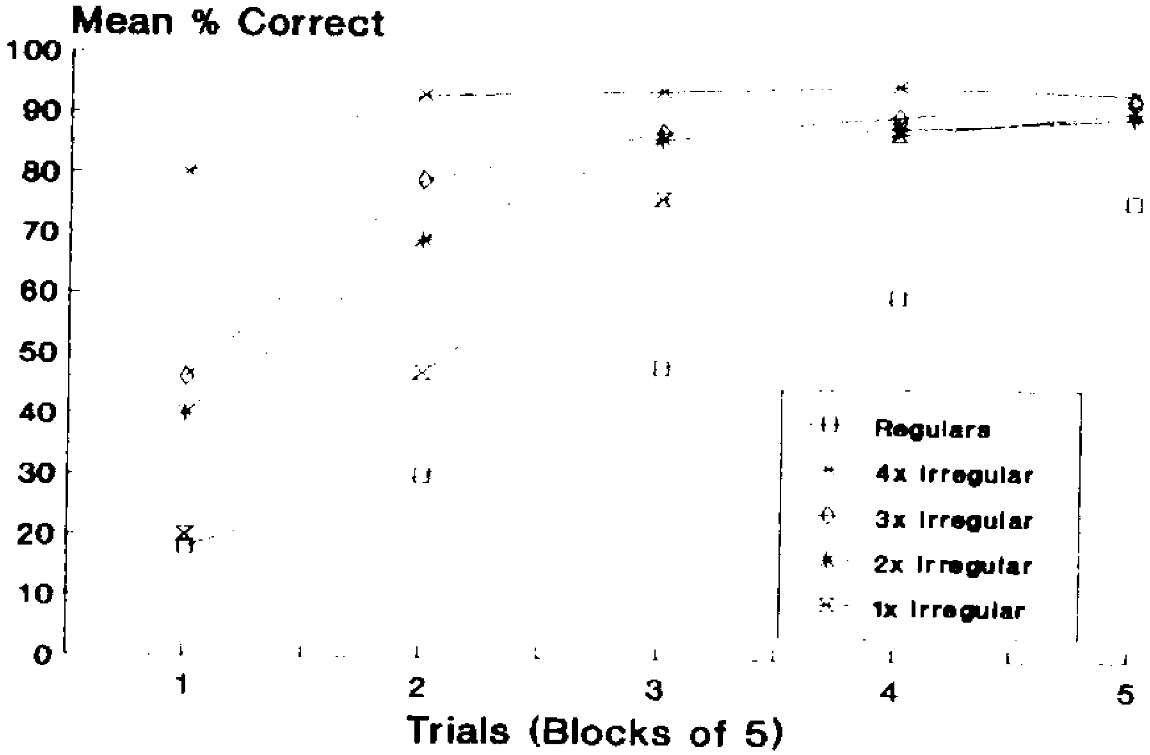


Figure 13. Palermo and Howe (1970): Mean percent correct/block of five trials for the regular pairs and each of the four different frequency irregular pairs.

likelihood of an error being an OR error increased[12].

Table 4 shows the relationship between OR errors and the frequency of presentation of the irregular pairs. Over all trials, and during each of the listed phases of acquisition, there is an inverse relationship between the likelihood of an OR error, and the frequency with which the irregulars were presented. As the frequency of presentation of the irregulars increased, the likelihood of an OR error decreased so that the irregulars presented four times per trial showed the fewest OR errors and the irregulars presented once showed the most.

The reanalysis of Palermo and Howe's experiment, conducted as part of this work, revealed several additional details which will be used to compare P&H's experiment to the SDM simulations detailed below. First, the average number of trials to a criterion level of one errorless trial was calculated for each type of irregular and is compared to

---

[12] There is a discrepancy between P&H's published results and the findings of the reanalysis performed as part of this work. P&H stated that there were a total of 158 OR errors on all trials, and 64% of these occurred after criterion was reached on the irregulars. This reanalysis identified 99 more OR errors than P&H reported, and most of the OR errors were found to occur BEFORE criterion was reached on the irregulars. At first glance the disparity in the distribution of OR errors before and after criterion seems glaring. P&H stated that most of the errors occurred after criterion, while this reanalysis indicates that most occurred before. However, this reanalysis also found that after criterion was reached on the irregulars, the majority of all errors are accounted for by OR errors. Perhaps this is what P&H were actually referring to, and the language used to convey this was simply misleading.

Table 4

Overregularization (OR) Errors Based on Frequency of Presentation of Irregular Pairs:Palermo and Howe (1970)[13]

	Frequency of Presentation of Irregular Pair (per trial)			
	4X	3X	2X	1X
# OR errors on all trials	10.75	26.33	35	65
# OR errors BEFORE Irregular Criterion[14]	6	16.67	21	42
# OR errors AFTER Irregular Criterion	4.75	9.67	14	23
# OR errors AFTER Regular Criterion[15]	0.75	0.67	1.5	2

---

[13] The number of errors indicates the ratio of the number of errors to the frequency of presentation per trial.

[14] The irregular criterion represents the first errorless trial on ALL irregular pairs.

[15] The regular criterion represents the first errorless trial on ALL regular pairs.

the regulars and all pairs in Figure 14. Criterion was reached first by the irregulars presented four times, followed by those presented three times, twice and once per trial. Second, the mean percent correct for regulars and all irregulars, and the mean percent overregularization errors were calculated for all subjects, and the results were plotted as a function of trial, with trial 0 equal to the point when a criterion of one errorless trial was reached on the irregulars (Figure 15). This graph shows that P&H's subjects exhibited U-shaped learning on the irregular pairs (albeit a shallow U), accompanied by overregularization errors, while the regulars showed a steady increase in performance. Finally, similar graphs were produced for each type of irregular, based on frequency of presentation (see Figures 16, 17, 18 and 19). A comparison of figures 16 through 19 reveals that as the frequency of presentation of the irregulars decreases, U-shaped learning becomes more pronounced and the percentage of OR errors increases.

2. SDM Simulations of Palermo and Howe (1970)

a. Method

As in the SDM simulations of the "rule of 78", the SDM simulations of P&H stored the data vector representing the correct response at the vector address of the stimulus. Once again, sixteen-bit vectors were used to represent both

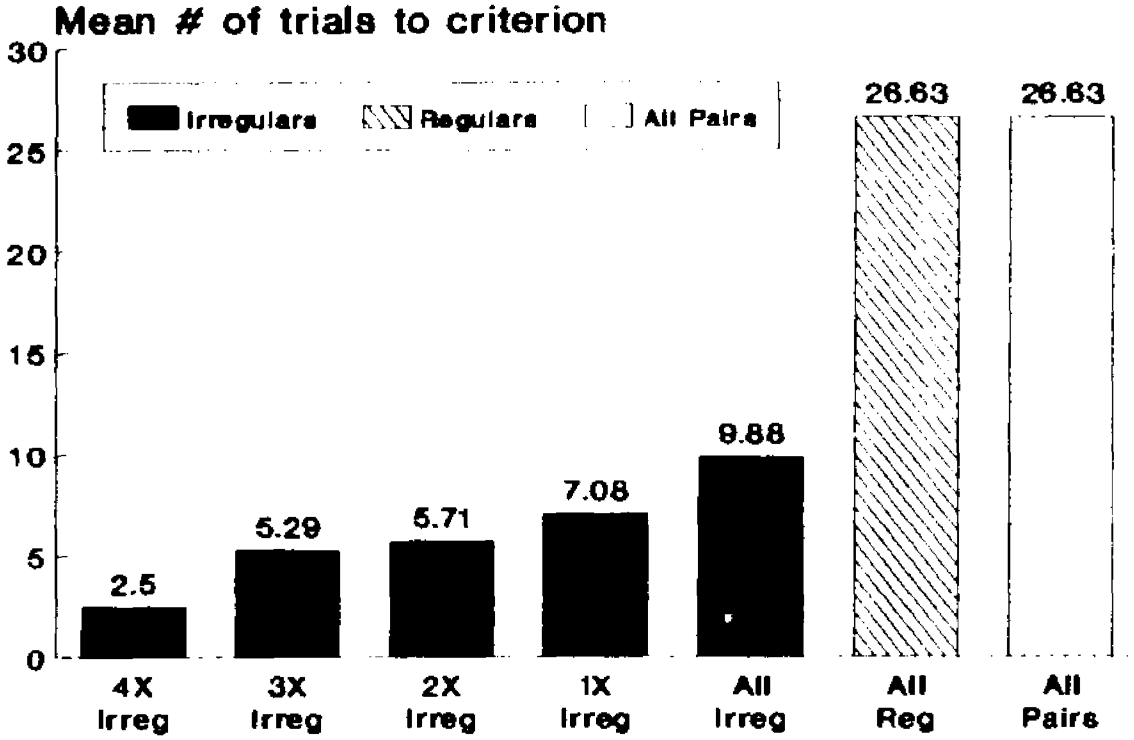


Figure 14. Palermo and Howe (1970): Mean number of trials to a criterion of one errorless trial.

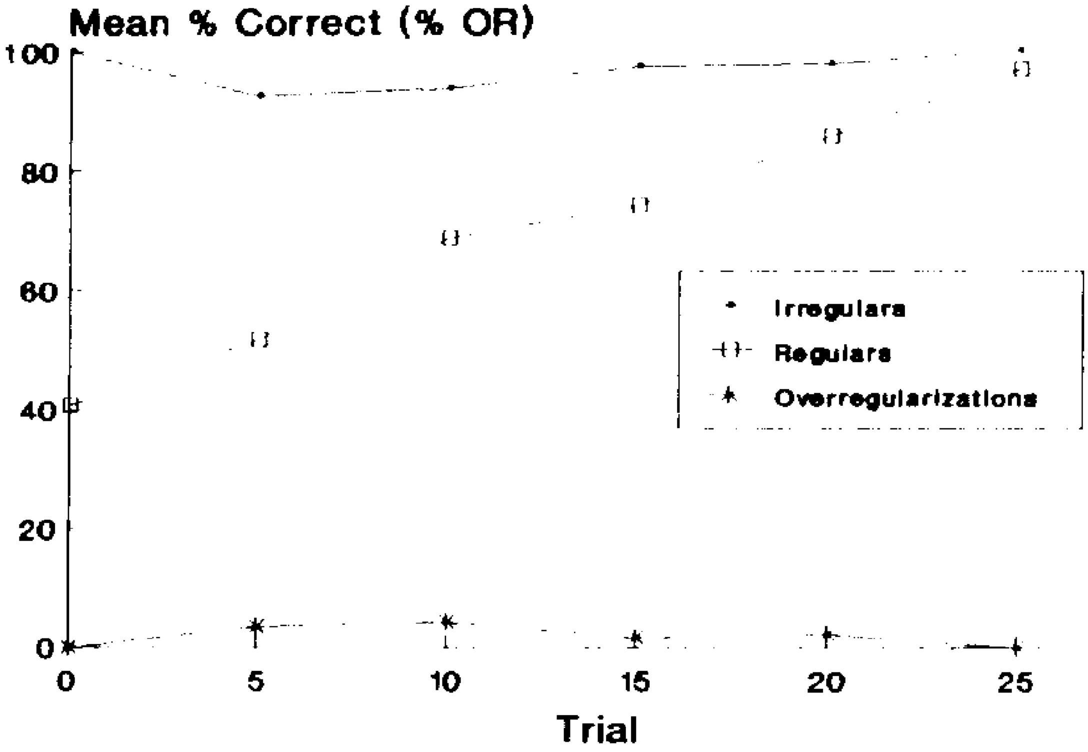


Figure 15. Palermo and Howe (1970): Mean percent correct/trial for regular and irregular pairs and mean percent OR errors/trial for irregular pairs, beginning with the trial during which each subject first reached 100% correct on the irregular pairs (trial C).

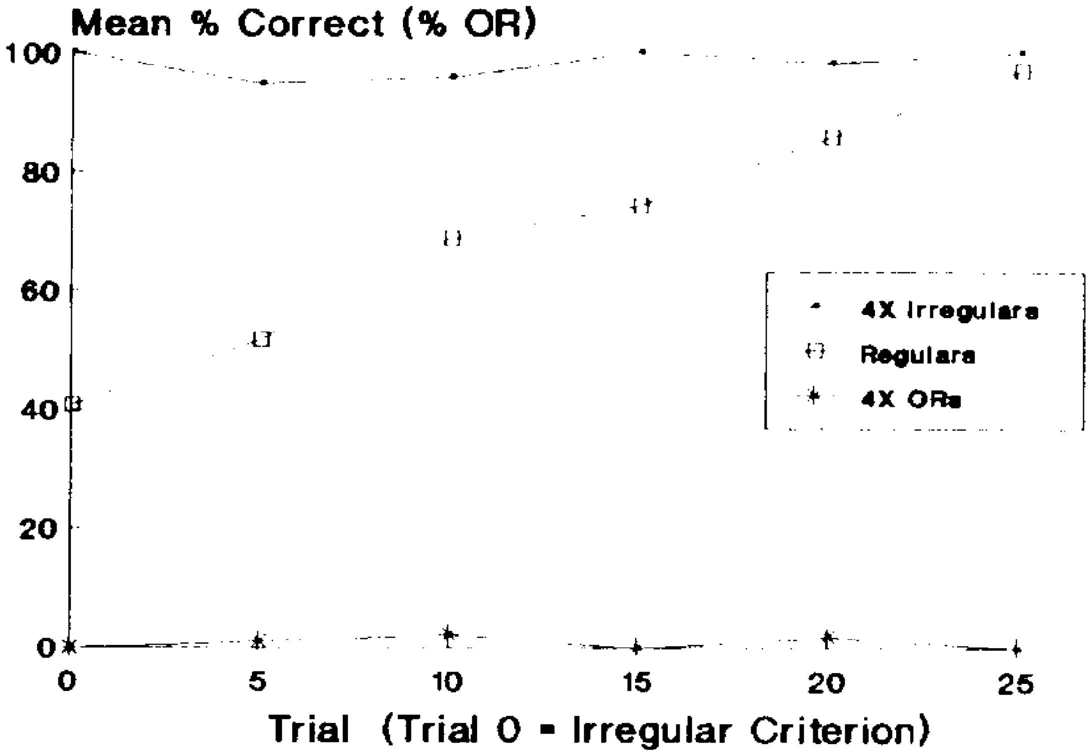


Figure 16. Palermo and Howe (1970): Mean percent correct/trial and mean percent OR errors/trial for the irregular pair presented 4X/trial, beginning with the trial during which each subject first reached 100% correct on all irregular pairs (trial 0).

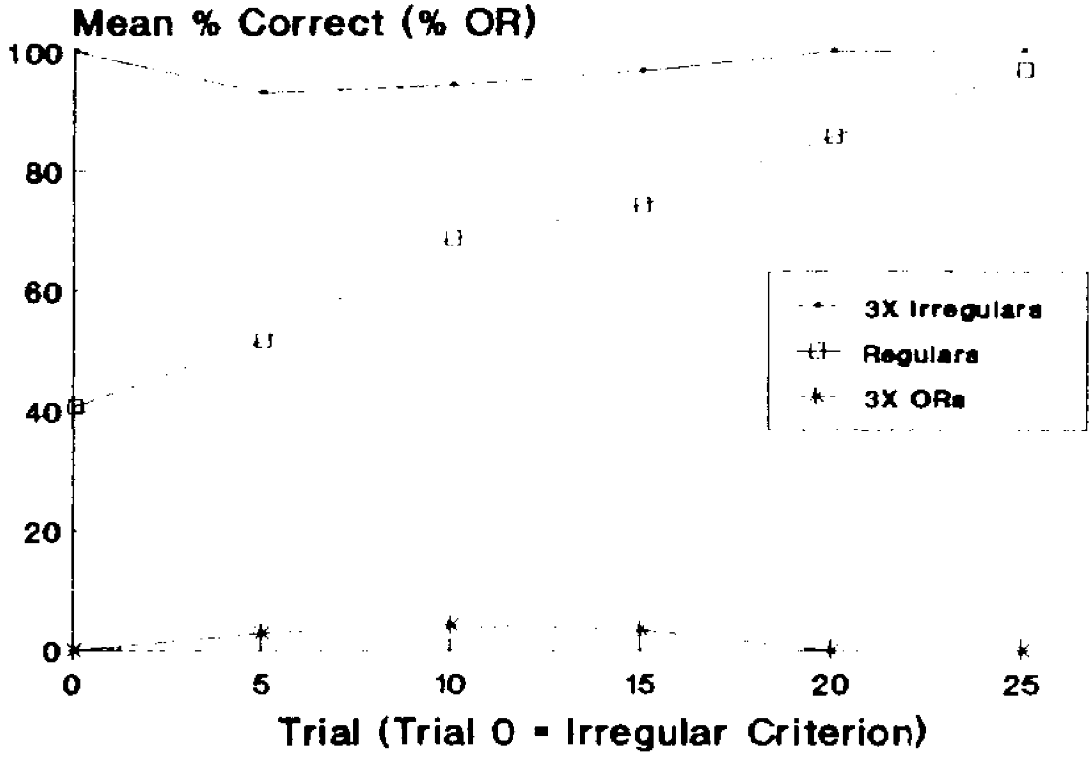


Figure 17. Palermo and Howe (1970): Mean percent correct/trial and mean percent OR errors/trial for the irregular pair presented 3X/trial, beginning with the trial during which each subject first reached 100% correct on all irregular pairs (trial 0).

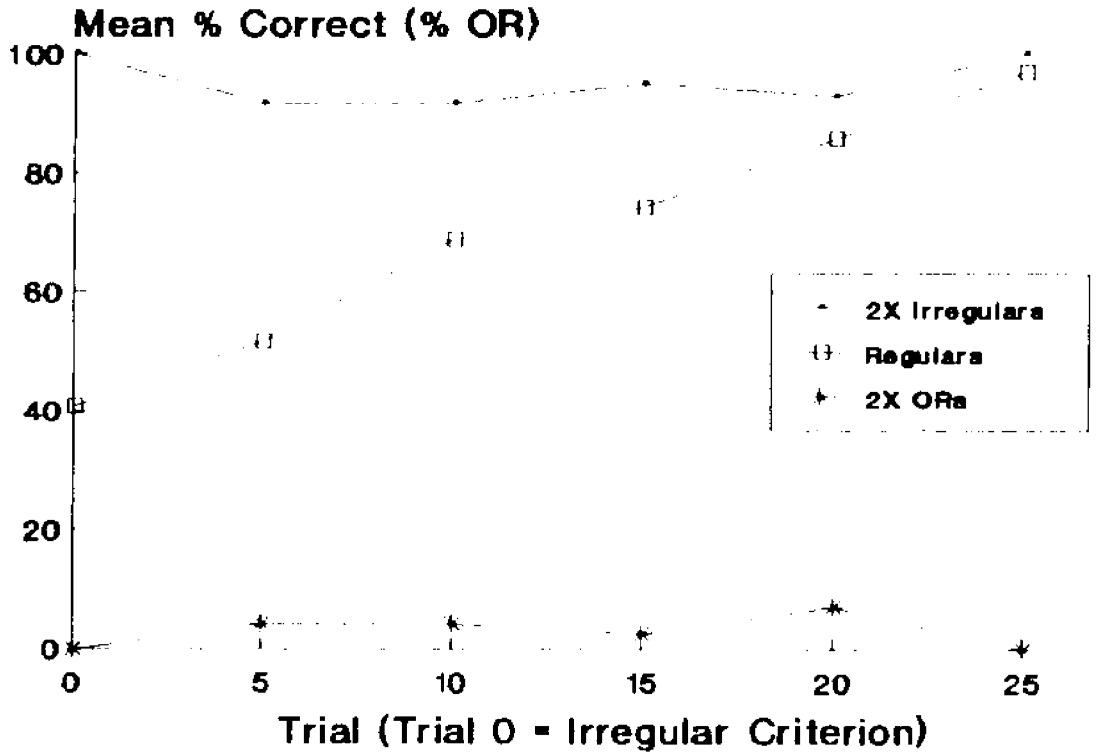


Figure 18. Palermo and Howe (1970): Mean percent correct/trial and mean percent OR errors/trial for the irregular pair presented 2X/trial, beginning with the trial during which each subject first reached 100% correct on all irregular pairs (trial 0).

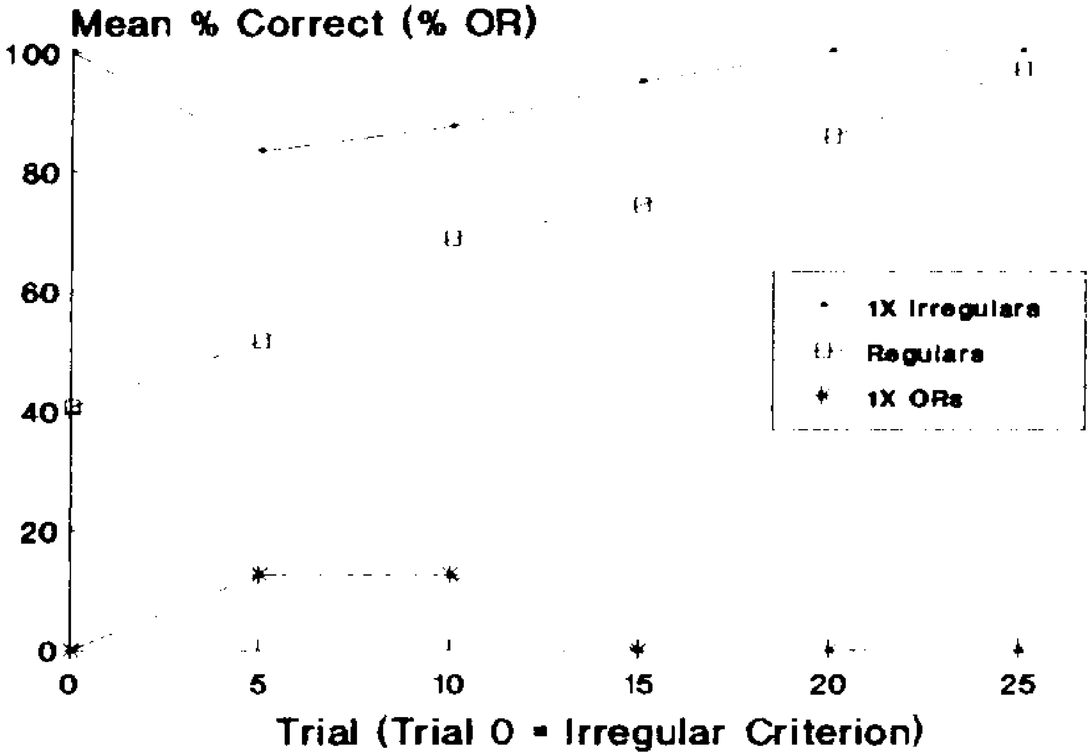


Figure 19. Palermo and Howe (1970): Mean percent correct/trial and mean percent OR errors/trial for the irregular pair presented 1X/trial, beginning with the trial during which each subject first reached 100% correct on all irregular pairs (trial 0).

the stimuli and responses. Prior to the addition of noise, the letter responses used in P&H's experiment were coded with non-overlapping bits in the data vectors. Table 5A shows the 16-bit vectors for each of the responses. Each response was represented by two unique bits in the vector so that the hamming distance between any two responses was 4 bits. Since 16-bit vectors were used, and the P&H stimuli were composed of two digits each, the first eight bits in the 16 bit address vectors represented the first digit of the stimulus, and the last eight bits were used to code the second digit of the stimulus. Since the digits 1 through 9 were used in P&H's stimuli, it was impossible to use two bits to code each digit while at the same time coding them in a non-overlapping manner[16]. However, a coding system was devised to maximize the hamming distance between the digits, given the constraints of P&H's design and the program's capabilities. The digits 1 and 2 were coded using only one bit, while the seven other digits (3-9) were coded using two bits. Table 5B shows the address bit-vectors for each digit. This coding system resulted in a mean hamming distance of 3 bits between any two digits with a range of 2 - 4 bits. Figure 20 shows two examples of the graphical representation of the stimulus and response vectors used in

---

[16] Coding items in a non-overlapping manner maximizes the hamming distance between items in the memory space. This is considered desirable in that non-overlapping items will not "share" as many hard locations and will therefore be easily distinguishable as unique items.

Table 5

SDM Simulations of Palermo & Howe (1970)

A. 16-Bit Vectors used to represent RESPONSES

RESPONSE		BIT-VECTOR
GROUPS A1 & B1	GROUPS A2 & B2	
A	A	1100000000000000
C	T	0011000000000000
I	I	0000110000000000
F	P	0000001100000000
O	O	0000000011000000
G	Q	0000000000110000
U	R	0000000000001100

B. 8-Bit Vectors used to represent STIMULUS DIGITS

STIMULUS DIGIT	BIT-VECTOR
1	00000001
2	00000010
3	01001000
4	00010100
5	10100000
6	00001100
7	00011000
8	01100000
9	11000000

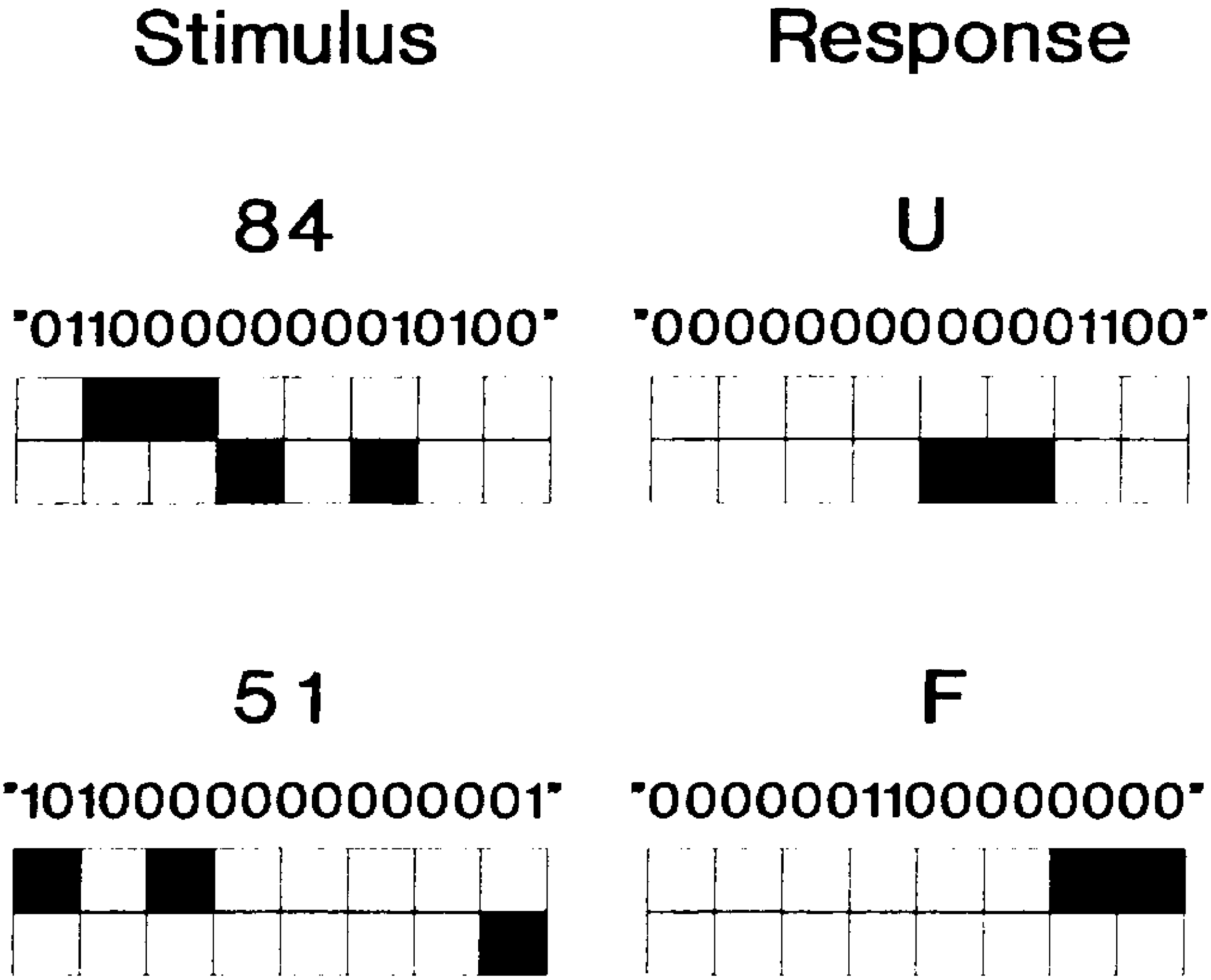


Figure 20. Graphical representation of the stimulus - response pairs used by the program in the SDM simulations of P&H(70). Two rows of 8 boxes were used to represent the 16 bit vectors. For the stimuli, the top row of eight boxes (and the first eight positions in the bit vector) represent the first digit of the stimulus, and the second row of boxes (the last eight positions in the bit vector) represent the second digit of the stimulus. All digits in the stimuli were represented by two bits, except for the digits "1" and "2", which were each represented by one bit. Each response was represented by two bits. Each box represents one bit. A black box represents a "1" in the bit vector and indicates that the bit is turned on. A white box represents a "0" in the vector and indicates that the bit is off.

the SDM simulations of P&H's experiment.

Five thousand "hard" memory locations were used in all the simulations reported here. During training, responses were stored at all addresses within a radius of 2 from the designated stimulus address, and during testing the responses were read from all addresses within a radius of 2 from the designated stimulus address. As in the SDM simulations of the "rule of 78", a radius of 2 was chosen to include 1/1000th of the total memory space. Twelve simulations were run. Each simulation consisted of five hundred trials. Each trial included a training session followed immediately by a testing session. Since P&H presented a different stimulus-response set to each of their four subject groups (A1, A2, B1 & B2), three simulations were run using each of these four different stimulus-response sets (see Table 3). Therefore during training in the Group A1 simulations, the presentation of the irregular pairs entailed presenting "84-U" four times, "46-I" three times, "57-O" twice, and "71-A" once per trial, while in the Group B2 simulations, "93-Q" was presented four times, "42-T" three times, "85-P" twice, and "26-R" once per trial. Within each group of three simulations, only the seed used to initialize the memory space differed. A separate computer program was designed to randomly generate ten sets of 12 regular pairs for each group of three simulations, based on the same set of constraints imposed by P&H (see section II.B.1). All 3 simulations within each group used the same ten sets of regulars, which were each presented every ten trials throughout each simulation. Therefore, a total of 120 regular pairs were presented fifty times each to the SDM program over the course of each simulation. During training, each regular pair was presented once per trial. During each testing trial, the data contained in each of the 22 stimulus addresses (12 regular and the 3 different irregulars, presented in the same frequency as during training for a total of 10 irregulars per trial) were read once. During both training and testing trials the noise level added to each stimulus address vector was initially set to 0.15. After every five trials this noise level was reduced by 0.01 until, during training, the level reached 0.02, and during testing, the noise level reached 0. Therefore, during training, trials 66 through 500 had a noise level of 0.02, and during testing, trials 76 through 500 had no noise added to the stimulus address vectors. Table 6 summarizes the parameters used in the 12 simulations of P&H's experiment.

b. Results

Eight of the 12 simulations reached 100% correct on all pairs at some point during the simulation and all of the simulations reached 100% correct on the irregulars. As in P&H, the point at which the regulars reached 100% correct in

Table 6

Parameters for the SDM Simulations of Palermo & Howe's Experiment

Vector length	16 bits
Number of hard memory locations	5000
Seed	randomly generated by the program
Number of trials per simulation	500
Stimulus-Response Sets	3 Simulations = P&H Group A1 3 Simulations = P&H Group B1 3 Simulations = P&H Group A2 3 Simulations = P&H Group B2
Training:	
Stimulus-Response Presentations (storing response at stimulus address)	10 sets of 12 different regular pairs. Each set of regulars presented once every 10 trials. Each regular presented 1X/trial. 1 irregular pair presented 4x/trial 1 irregular pair presented 3x/trial 1 irregular pair presented 2x/trial 1 irregular pair presented 1x/trial
Radius	2
Noise added to stimulus address vectors	Trial 1: Noise = 0.15, and is reduced by 0.01 after every 5 trials until Noise = 0.02 (trials 66-500)
Testing:	
Response Readings (retrieving response from stimulus address)	Same as training
Radius	Same as training
Noise added to stimulus address vectors	Trial 1: Noise = 0.15, and is reduced by 0.01 after every 5 trials until Noise = 0.00 (trials 76-500)

the simulations was also the point at which 100% correct was attained for all pairs, since the irregulars stabilized at 100% relatively early in each simulation. The mean trial at which the irregulars stabilized at 100% for the 12 simulations was 71.6, with a range of 68-75 trials. For the 8 simulations which reached 100% on the regulars, the mean number of trials to a criterion of one errorless trial for the entire list was 320, while the mean number of trials to the same criterion for the irregular cases (for the same 8 simulations) was 29[17]. None of the simulations reached criterion on the regulars before the irregulars and, for the 8 simulations which reached 100% on all pairs, there was an average of 291 trials between the points at which the two criteria were reached (range = 197 - 436 trials)[18]. As illustrated in Figure 21, except for the first 10 trials, when the irregulars presented one time per trial were identified correctly 21.67% of the time and the regulars were correctly retrieved 23.68% of time, performance on the

---

[17] Four of the simulations, (three from Group A2 and one from Group B2) reached 91.7% (one wrong) at some point during the simulation. If this point is considered to be the point at which these simulations reached criterion, the mean number of trials to criterion for the entire list, for all 12 simulations, was 286, while the mean number of trials to criterion for the irregulars was 30.

[18] If the point at which the above mentioned four simulations reached 91.7% correct on all pairs is taken to be the point at which these simulations reached criterion, the average number of trials between the points at which criterion was reached for all pairs and for the irregulars was 256 for all 12 simulations, with a range of 120 - 436.

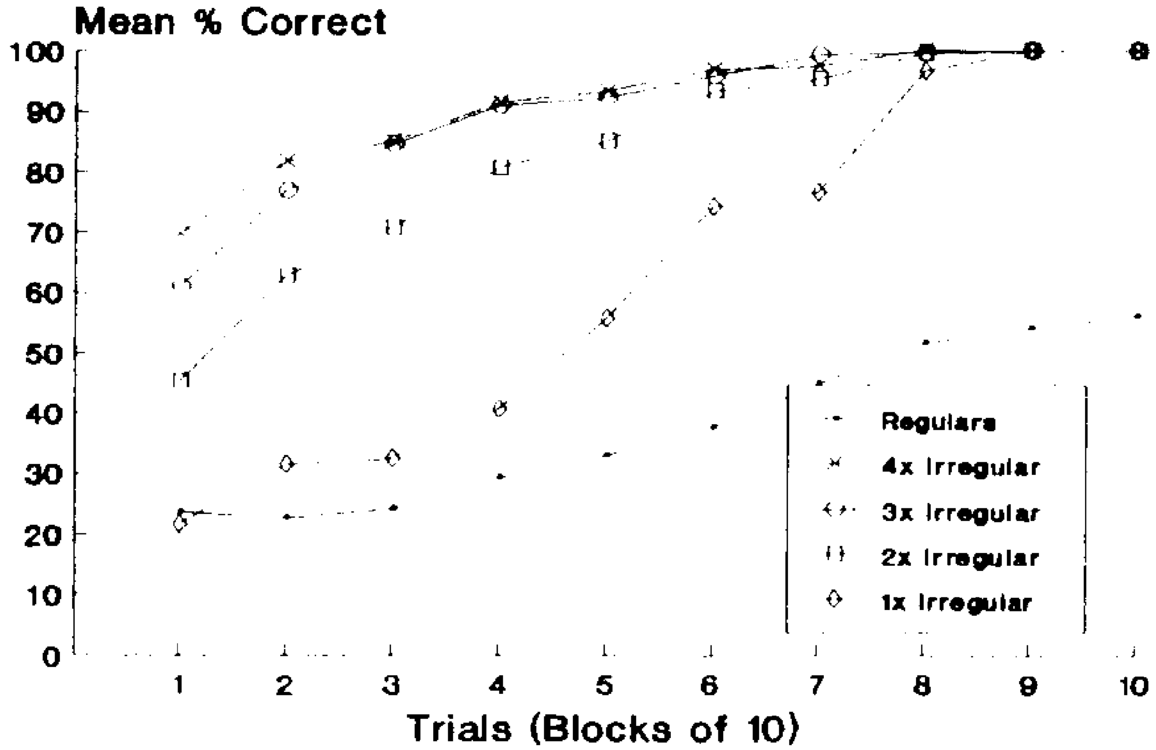


Figure 21. SDM simulations of P&H(70): Mean percent correct/block of ten trials for the regular pairs and each of the four different frequency irregular pairs.

regular forms was consistently below that of all the irregulars. Figure 21 also shows an orderly increase in acquisition rates based on the number of presentations per trial for each irregular[19]. During the first 30 trials, the irregular presented four times per trial was learned slightly faster than the irregular presented three times per trial, at which time the acquisition curves for the two types of irregulars are barely distinguishable (although except for trials 61-70, the percent correct for the 4X irregular is insignificantly higher). The irregular presented twice per trial was learned next and the irregular presented once per trial showed the slowest acquisition rate among the irregulars.

A total of 82 overregularization (OR) errors occurred throughout the twelve simulations. The majority of these OR errors (69 or 85% of all OR errors) occurred before criterion was reached on the irregulars and 12 (15% of all OR errors) occurred after criterion was reached on the irregular pairs. OR errors accounted for 5% of all errors made on irregulars over all trials, 6.3% of all irregular errors before criterion was reached on the irregulars, and 2.4% of all irregular errors after criterion was reached on

---

[19] Figure 21 only shows the first 100 trials of the simulations. During the last 400 trials, all the irregulars remained at 100% correct, but the regulars continued to show steady improvement and by trials 490-500, on average, 85% of all regular responses in all 12 simulations were correctly retrieved.

the irregulars. No OR errors (or, for that matter, any kind of irregular errors) occurred after a criterion of one errorless trial on the regular pairs was achieved (or 91.7% correct on the previously mentioned simulations which did not reach 100% on the regulars). Other types of errors included omissions (a response consisting of all zeros was retrieved from the designated stimulus address), partial retrievals, in which only one bit of the two bits used to represent the response was retrieved, and finally completely incorrect responses in which a response other than the correct response was retrieved from the stimulus address.

Table 7 indicates the relationship between OR errors and the frequency of presentation of the irregular pairs in the simulations. Over all trials and after criterion was reached on the irregulars, there is an inverse relationship between the likelihood of an OR error and the frequency with which the irregulars were presented. That is, as the frequency of presentation of the irregulars increased, the likelihood of an OR error decreased. Except for the 3X ORs which are higher than the 2X ORs, this also holds for the trials before criterion was reached on the irregulars. Therefore, in all cases, the likelihood of an OR error is greatest for the irregular presented once per trial and least for the irregular presented four times.

Figure 22 compares the average number of trials to a criterion of one errorless trial for each type of irregular,

Table 7

Overregularization (OR) Errors Based on Frequency of Presentation of Irregular Pairs - SDM Simulations of P&H[20]

	Frequency of Presentation of Irregular Pair (per trial)			
	4X	3X	2X	1X
# OR errors on all trials	4.75	7	7.5	26
# OR errors BEFORE Irregular Criterion[21]	4.25	6.33	6	21
# OR errors AFTER Irregular Criterion	0.5	0.67	1.5	5
# OR errors AFTER Regular Criterion[22]	0	0	0	0

---

[20] The number of errors indicates the ratio of the numberof errors to the frequency of presentation per trial.

[21] The irregular criterion represents the first errorless trial on ALL irregular pairs.

[22] The regular criterion represents the first errorless trial on ALL regular pairs (or for those 4 simulations which did not reach 100% on the irregulars, the first trial at 91.7% correct).

all irregulars, the regulars and all pairs[23]. A comparison among the irregulars of different frequencies reveals that criterion was reached first by the irregular presented four times per trial and last by the irregular presented once per trial, although the difference between the two is rather modest (less than one trial). Figure 23 shows the mean results of the 12 simulations beginning at the point each simulation reached criterion on the irregulars (labelled trial 0) and continuing through one hundred subsequent trials. While the regulars show a steady increase in performance, irregular performance drops below perfect in the 10 trials following criterion and then returns to perfect performance over the next 50 trials, resulting in a U-shaped acquisition curve. Low levels of overregularization (0.8% - 3.3%) accompany this period of U-shaped learning. In the 350+ trials following the trials depicted in Figure 23, the irregulars remain at 100% correct, while the regular responses continue to be identified more correctly as the simulations progress, until by the end of the simulations they reach an average of 85% correct. Figures 24 through 27 show similar graphs for each type of irregular, based on frequency of presentation. A

---

[23] The values given for the number of trials to criterion for the regular pairs and all pairs include only the 8 simulations which reached 100% correct on the regulars at some point in the simulation. The values for the number of trials to criterion for all the irregulars include all 12 simulations.

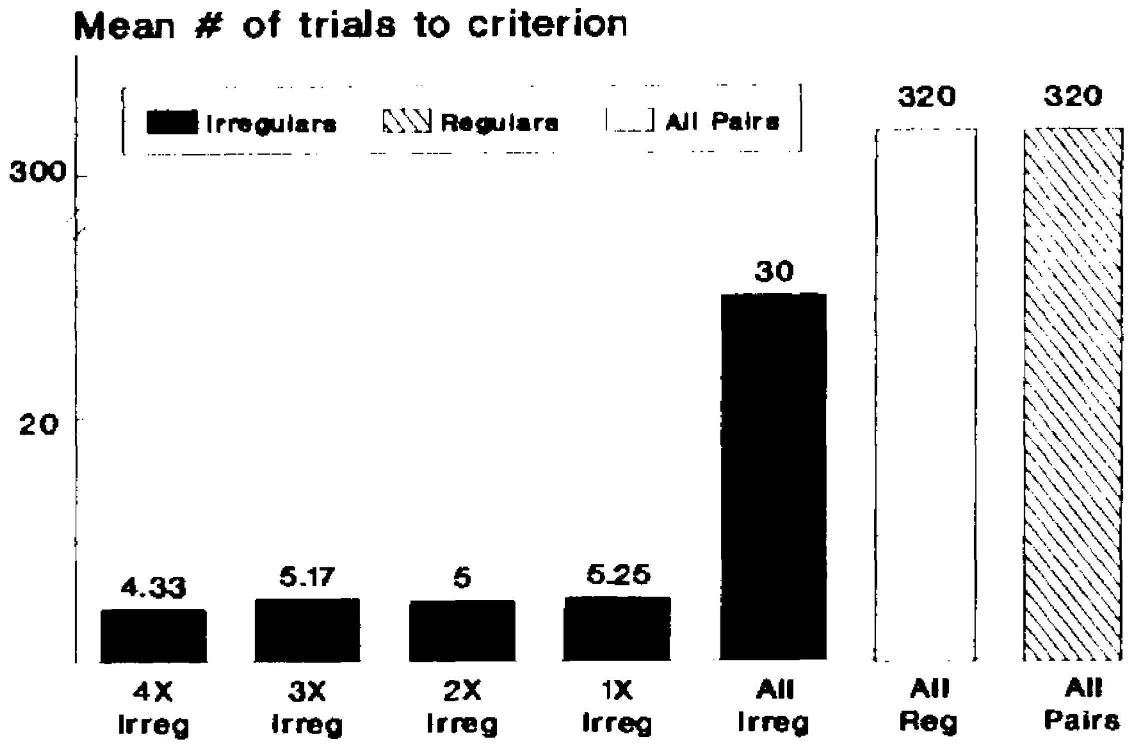


Figure 22. SDM simulations of P&H(70): Mean number of trials to a criterion of one errorless trial.

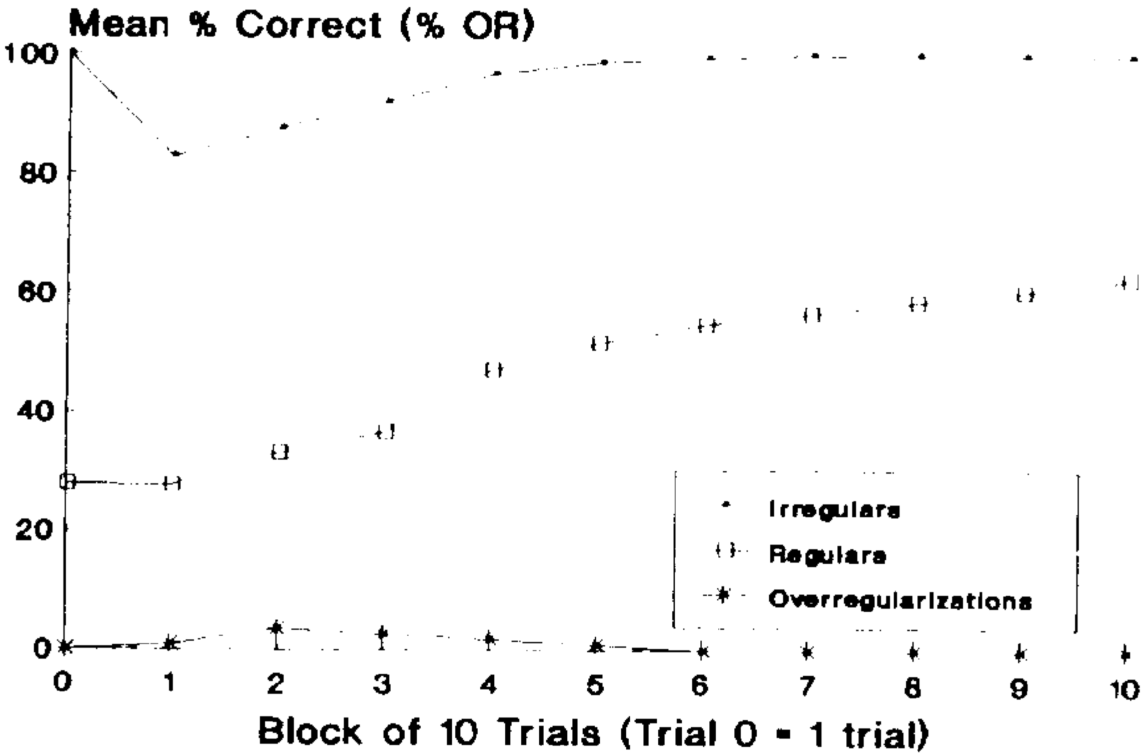


Figure 23. SDM simulations of P&H(70): Mean percent correct/block of 10 trials for regular and irregular pairs and mean percent OR errors/block of 10 trials for irregular pairs, beginning with the trial during which each simulation first reached 100% correct on the irregular pairs (trial 0).

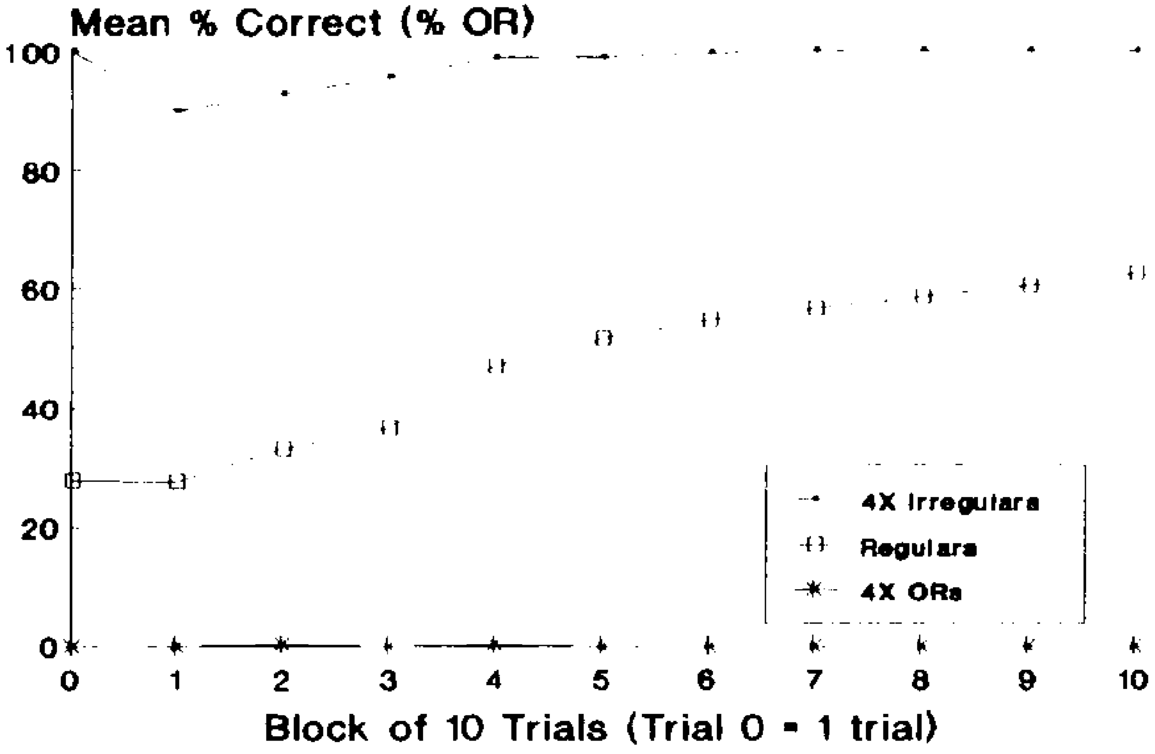


Figure 24. SDM simulations of P&H(70): Mean percent correct/block of 10 trials and mean percent OR errors/block of 10 trials for the irregular pair presented 4X/trial, beginning with the trial during which each simulation first reached 100% correct on all irregular pairs (trial 0).

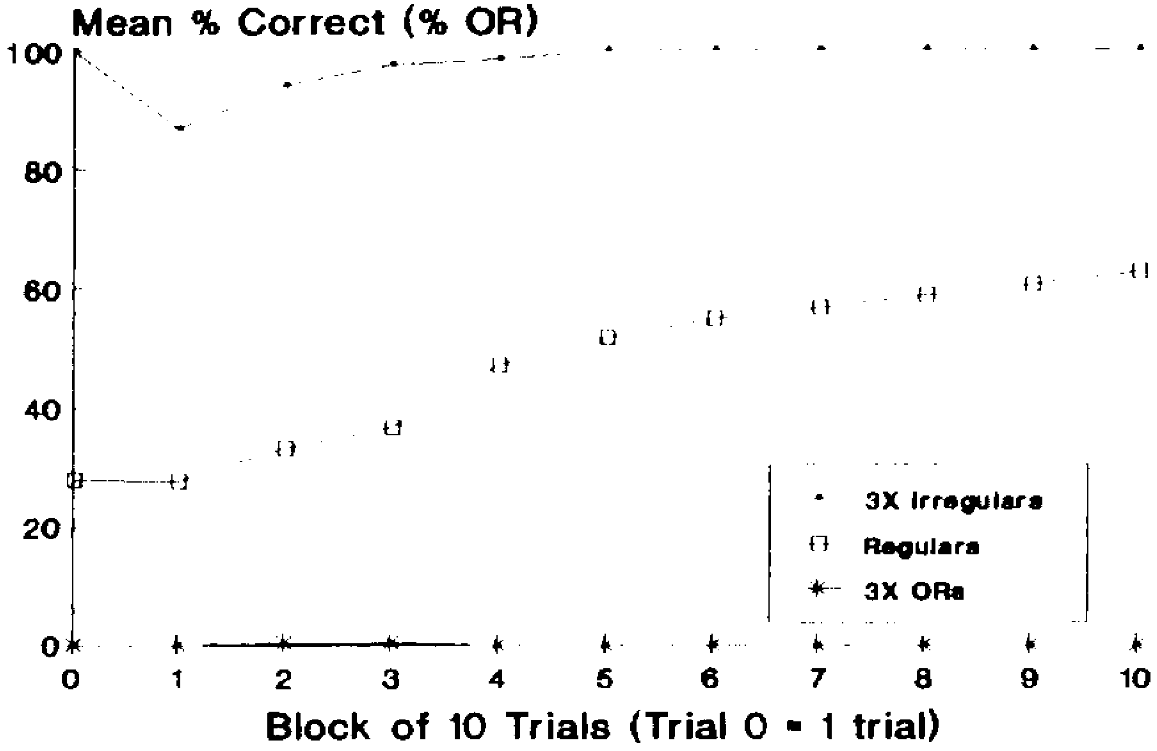


Figure 25. SDM simulations of P&H(70): Mean percent correct/block of 10 trials and mean percent OR errors/block of 10 trials for the irregular pair presented 3X/trial, beginning with the trial during which each simulation first reached 100% correct on all irregular pairs (trial 0).

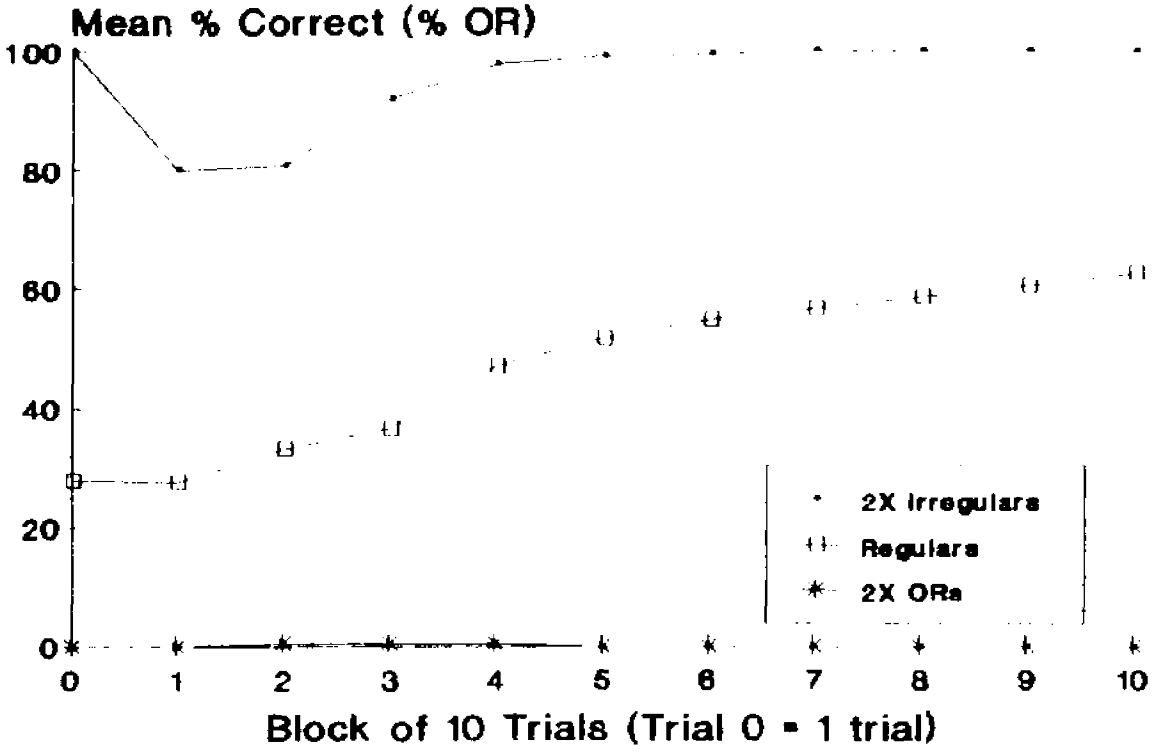


Figure 26. SDM simulations of P&H(70): Mean percent correct/block of 10 trials and mean percent OR errors/block of 10 trials for the irregular pair presented 2X/trial, beginning with the trial during which each simulation first reached 100% correct on all irregular pairs (trial 0).

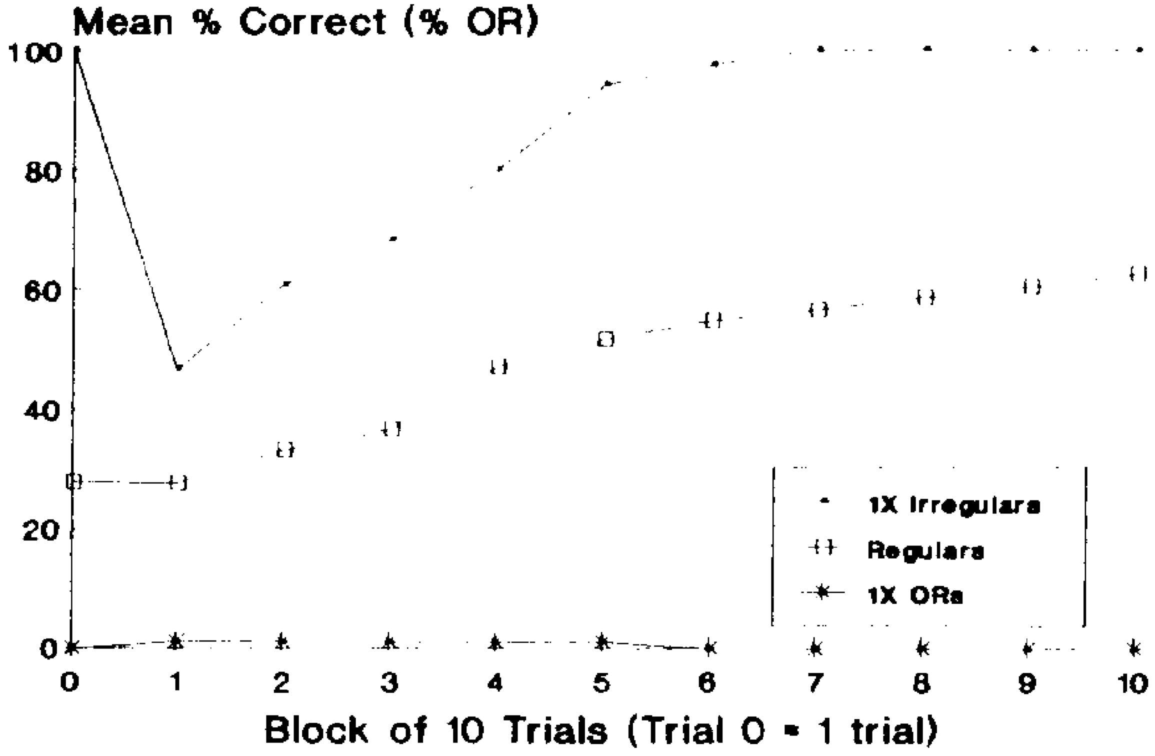


Figure 27. SDM simulations of P&H(70): Mean percent correct/block of 10 trials and mean percent OR errors/block of 10 trials for the irregular pair presented 1X/trial, beginning with the trial during which each simulation first reached 100% correct on all irregular pairs (trial 0).

comparison of Figure 24 (the 4X irregular) and Figure 27 (the 1X irregular) shows the most glaring difference in acquisition based on frequency. The irregular presented 4X per trial only shows a shallow U (a dip in performance of only 10% at its lowest point) and very low OR rates (0.2% OR for trials 11 - 30 following criterion and 0% OR during all other trials). In contrast, the irregular presented 1X per trial shows a drop in performance of 53% at the lowest point in the U, and a higher (albeit still low) level of overregularization for a longer period of time (1% OR for trials 1-50 following criterion). The irregular presented 2X per trial (Figure 26) shows a U much shallower than the irregular presented once per trial, but also distinguishably deeper than the U-shaped learning exhibited by the irregulars presented four times per trial and 3X per trial.

3. Discussion

Palermo and Howe (1970) demonstrated that U-shaped learning and overregularization are not phenomena exclusive to the language acquisition domain, but instead, can also occur in a laboratory setting in which adults are presented with the task of learning a simple system that can be described as containing rules and exceptions. The goal of the SDM simulations reported here was not only to see if SDM could learn the same system that P&H's subjects learned, but to see if the model would exhibit some of the same characteristics of the learning process, including U-shaped acquisition and overregularization. With this goal in mind, various aspects of P&H's experiment will be compared to the SDM simulations in order to lend further support to the notion that SDM is a viable model of human memory.

Before comparing the results of these simulations to P&H's subjects, an important distinction must be made between the two situations. Since the details of P&H's methods and results were made available to me, it was possible to mimic the situation encountered by P&H's subjects in the simulations. That is, the same irregulars were presented, and the regulars were generated in the same way. However, there are various differences between humans and computers and one of these is critical in the evaluation of these simulations. The Hamming distances between numbers and letters in these simulations were designed to be basically equidistant, thereby minimizing the associations between any two letters or numbers. In contrast, it has been established that humans often make use of associations between letters and numbers. In other words, for humans, letters and numbers are not equidistant. Conrad (1963, 1964) has shown that some letters are psychologically more closely related than others based on their phonemic and featural characteristics. For example, in memory humans often confuse the letters B and P, which both look and sound alike, but they rarely confuse X and P, which share few, if any phonemic characteristics. In addition, Miller (1956) in his often cited discussion of the "magical number 7 ± 2", has shown that people often chunk items based on meaning, in order to hold more information in memory. Therefore, the subjects in P&H's experiment have an advantage over the simulations because of their previous experience with both numbers and letters. Thus, although the stimulus - response sets used in P&H's experiment and these simulations appear identical on the surface, in reality, they are fundamentally different and any comparison of the results must take this difference into consideration.

Despite this underlying distinction, the SDM simulations were able to capture many of the characteristics of the acquisition process revealed in P&H's experiment. First, based on a comparison of the number of trials to a criterion of one errorless trial, all the irregulars were acquired before all the regulars (see Figures 14 and 22). Although the simulations took much longer to reach criterion on the regulars than P&H's subjects did, this might be due to the prior experience and semantic associations employed by humans when faced with the task of learning new associations. In P&H's experiment, the distinction between regular and irregular responses was correlated with a vowel/consonant distinction. For example, in P&H's subject groups A1 and A2, all irregulars required vowel responses, while all correct regular responses were consonants. P&H's

subjects might have used their previous experience with this distinction to assign more structure to their responses than was possible in the simulations. Aside from this departure, the acquisition process appears remarkably similar. Both P&H's subjects and the SDM simulations learned the irregulars long before the regulars. In both cases, when criterion was reached on the regulars, the whole list also appeared to be mastered. After the regular criterion was reached, P&H's subjects were almost perfect on the irregulars (99% correct) while still making a number of errors on the regulars (93% correct). Similarly, the program perfectly identified the irregulars during the 180 trials after criterion was reached on the regulars but was only 85% correct on the regulars by trial 500. Therefore, in both situations, not only were the irregulars learned more rapidly, they were also learned better than the regulars. In addition, in both instances the irregular presented four times per trial reached criterion before all the other irregulars and the irregular presented once per trial took longest to reach criterion.

Although the difference between the number of trials to criterion between the 4X and 1X irregular in the simulations is small (less than 1 trial), if the number of trials to asymptotic performance at 100% correct is calculated for the 12 simulations, the difference becomes more obvious. On average, the irregular presented four times per trial took

59 trials to stabilize at 100% correct, while the irregular presented once per trial required 69 trials to consistently perform perfectly in the simulations. Since the irregular is presented four times per trial, there are four opportunities for mistakes per trial, while for the irregular presented once, there is only one opportunity to err. Therefore, the number of trials to a criterion of one errorless trial may be misleading, especially for irregulars presented with less frequency, and the number of trials to consistently perfect performance may be a better indicator of proficiency. This idea is further supported by the fact that, in the simulations, the number of trials to a criterion of one errorless trial was 5.17 for the irregular presented three times and only 5 for the irregular presented twice per trial. However, the number of trials to consistently perfect performance in the simulations was 54 for the 3X irregular and 60 for the 2X irregular.

Secondly, in both P&H's experiment and in these simulations, performance on the regulars was consistently below that of the irregulars (refer to Figures 13 and 21). The only exception to this occurred in the first 10 trials of the simulations during which performance on the regulars was insignificantly higher than that of the irregulars presented one time per trial (23.68% for the regulars vs. 21.67% for the 1X irregulars). Additionally, both P&H's subjects and the simulations displayed an orderly increase in irregular acquisition rates based on the frequency of presentation of the irregulars. That is,the irregular presented four times was learned before the irregular presented three times, followed by the irregular presented twice and finally by the irregular presented once per trial. In both cases, the most significant difference in acquisition rates was between the irregular presented four times per trial and the irregular presented once per trial. In the simulations, the 4X irregular and the 3X irregular showed the most similar patterns of acquisition, while in P&H's experiment, the irregular presented three times and the irregular presented twice showed the most similar learning curves. A possible explanation for this difference lies in the bit-vectors used to represent these particular irregulars. While the mean Hamming distance between the bit-vectors used to represent the 4X irregulars and the 3X irregulars was only 6 bits, the mean Hamming distance between the 3X irregulars and the 2X irregulars was 6.75 bits. In other words, the bit-vector representations of the 4X irregulars and the 3X irregulars were more similar (shared more bits) than those used to represent the 3X and 2X irregulars. Had the representations of the 3X and 2X irregulars been at least as similar as the 4X and 3X irregulars, perhaps the acquisition pattern would have looked more like that of P&H's subjects in this respect.

Third, and perhaps most importantly for the purposes of this work, in both P&H's experiment and in the SDM simulations, U-shaped learning is evident once criterion is reached on the irregulars (see Figures 15 and 23). In both cases the U is rather shallow. At their lowest points there was a decrease in performance of approximately 8% for P&H and 17% for the simulations, and both of these lows occurred within the first 10 trials following the irregular criterion. In addition, during this period of U-shaped acquisition, low levels of overregularization (less than 5% in both P&H's experiment and the simulations) were evident. The simulations extended well past the point at which the irregulars were consistently identified perfectly, while P&H's experiment was terminated before their subjects reached a prolonged period of accuracy on the irregulars. In all probability this, more than anything, accounts for the fact that the right hand arm of the U is not as evident in P&H's experiment as in the simulations. In both the human data and in the simulations, there appears to be a positive relationship between frequency of presentation and correct performance and an inverse relationship between frequency and overregularization rates (refer to Figures 16 through 19 and 24 through 27). Additionally, in both situations, the inverse relationship between frequency and OR is further borne out by the fact that as the frequency of the irregulars increased, the number of OR errors, adjusted for the number of presentations per trial, decreased (see

Tables 4 and 7).

Thus far, the similarities between P&H's subjects and the simulations have been stressed, since they include some of the most important characteristics of the learning process. However, there are a number of differences between the human data and the simulations which should be noted. First, all the subjects included in P&H's analysis reached criterion on the regulars (and therefore reached criterion on all pairs), while four of the simulations did not. However, it must be pointed out that P&H state that data from four additional subjects were discarded since these subjects failed to reach a criterion of one errorless trial by trial 35 (their cut-off point for the experiment). Therefore, the simulations, since they include instances in which the overall criterion is not met, may actually be more representative of the general population than the data presented by P&H.

Secondly, P&H observed an increase in performance to above chance levels on the regulars following irregular criterion, in addition to an increase in proportion of OR errors at this point. These effects, which were not seen in the simulations, are particularly important since P&H use them to support their two process learning theory. The two process learning strategy includes learning the irregular items by rote memorization while applying the rule to the regulars. According to P&H, the application of the rule is not seen as advantageous until the irregulars, which according to this theory must be memorized, are mastered. Once the rule is seen as advantageous, P&H claim that regular performance should rise above chance levels. Figure 28 shows that regular performance does indeed increase approximately 10% (from approximately 31% one trial before irregular criterion to 41% at trial 0). Presumably it is at this point that the irregulars are mastered. The evidence presented to support this theory is based on a chance level of 33% for regular responding. This level of chance assumes that their subjects knew that one of the three regular responses was correct, knew which responses were regular, and finally, that responding on regulars is not affected by irregular interference. This final assumption poses the greatest theoretical problem. Assuming no interference from the irregulars is only valid if two completely separate processes for irregulars and regulars are employed. Yet, only by making this assumption can P&H make the claim that regular responding is below chance before the irregular criterion, and above chance following it. If the assumption that there is no interference from the irregulars is thrown out, then chance regular responding must be based on all seven possible responses and falls to 14.3%. Without this assumption, P&H's subjects are above chance well before the irregular criterion is reached and P&H's argument for a two process theory is only supported by the fact that the

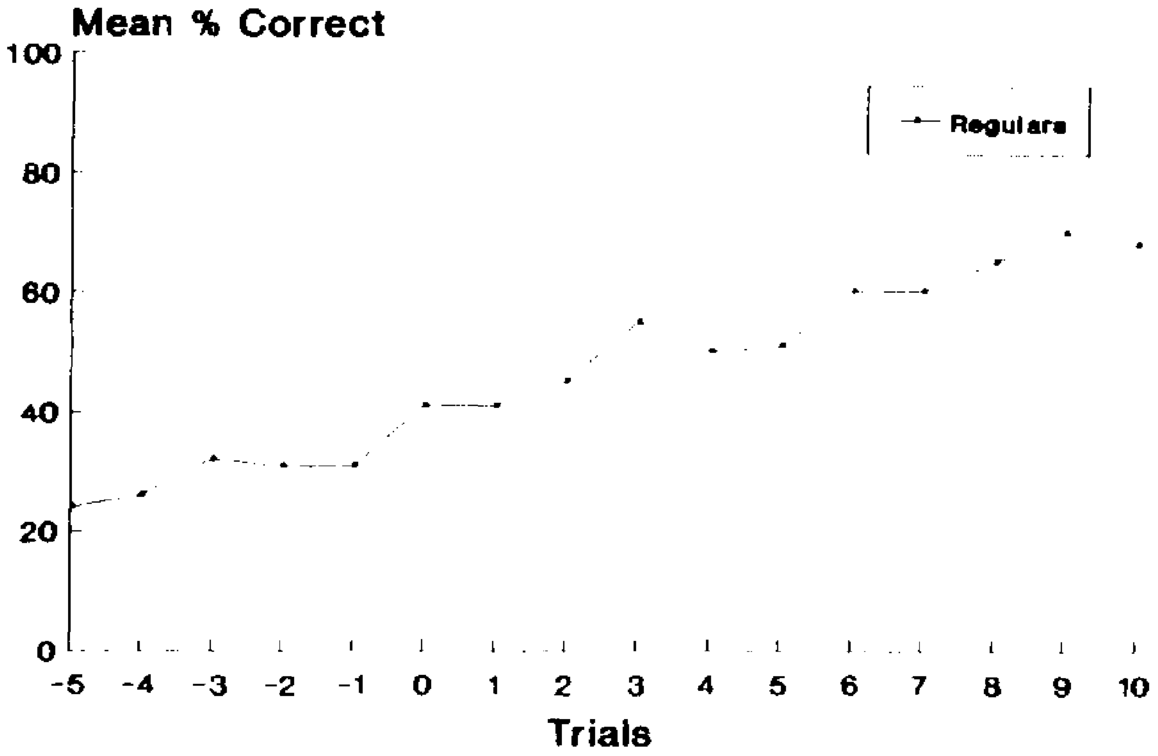


Figure 28. Palermo and Howe (1970): Mean percent correct/trial for the regular pairs. Trial 0 = the trial during which each subject first reached 100% correct on the irregular pairs.

proportion of OR errors increase after criterion is reached on the irregulars.

The claim is being made here that SDM can account for U-shaped learning and OR in a system that requires only one process to learn both regulars and irregulars, since all instances are learned in the same way, through distributed representation of the input and statistical reconstruction from multiple memory locations for the output. Therefore, interference from irregulars cannot be ruled out in SDM and chance levels of responding (whether it be for the regulars or the irregulars) is 14.3%. Not surprisingly, in the SDM simulations, just as in P&H, regular responding was above this chance level both before and after criterion was reached on the irregulars (see Figure 29).

However, in this work, SDM was not able to simulate the results achieved by P&H which indicates that as subjects became more proficient with the regulars, OR errors accounted for a higher percentage of the total errors. That is, after criterion was reached on the irregulars, OR levels accounted for 54% of all errors, whereas before this same criterion was reached, only 19% of the total errors were accounted for by OR errors. In contrast, in the simulations, OR errors accounted for only 2.4% of all errors after the irregular criterion, and 6.3% of all errors before the same criterion. Since this effect is one of P&H's main arguments for a two process theory, and SDM did not show

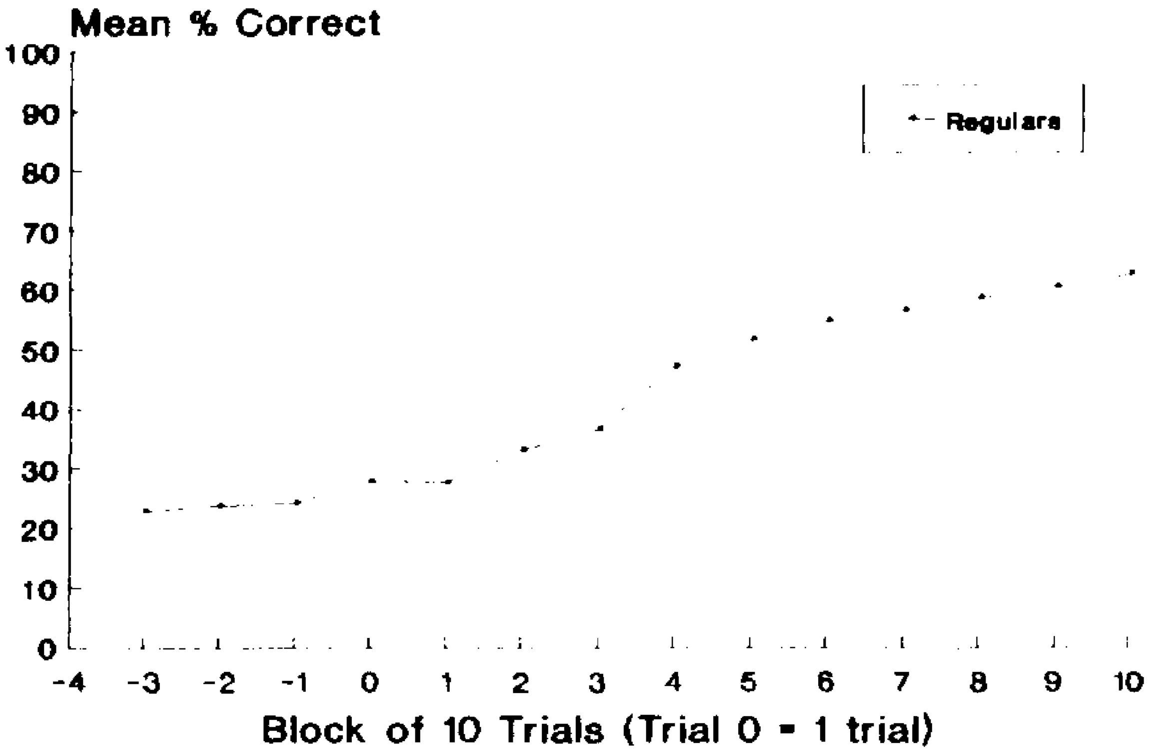


Figure 29. SDM simulations of P&H(70): Mean percent correct/block of 10 trials for the regular pairs. Trial 0 = the trial during which each simulation first reached 100% correct on the irregular pairs.

this effect in a one process system, the notion of SDM as a viable model of human memory must be addressed. The most reasonable explanation for this difference lies in the previously mentioned fundamental differences between these computer simulations and humans. These SDM simulations did not code the stimuli and responses in a way that would account for the types of associations that are commonly used by humans in memory tasks. In addition, computer simulations in general do not have the benefit of the prior experience that people do. One such association that could very well have been picked up by P&H's subjects is the vowel/consonant distinction, which, as previously mentioned, was correlated with the distinction between regular and irregular responses. Perhaps if the stimuli and responses used in these simulations had been coded to take into account this consonant/vowel distinction (or any of countless other associative patterns based on previous experience which P&H's subjects may have employed to help them learn this stimulus-response set), the simulations would have more closely matched P&H's results.

In addition, and more relevant in the case of the OR discrepancy, people often guess when they are unsure, but a model like SDM does not. If SDM cannot "find" a response from a summing of its memory locations, it generates an output consisting of all zeros. This output may be considered comparable to a human omission error, and these omissions accounted for approximately 50% of all responses in the simulations, but only 20% for P&H's subjects. If any of this difference in omission rates can be attributed to guessing on the part of P&H's subjects, and at least some of this guessing results in overregularization errors, then perhaps omission rates and OR rates are inversely related. In addition, P&H's subjects made even fewer omission errors in the five trials after criterion was reached on the irregulars (17%), while in the simulations the omission rate was just as high (50%) during the five trials after irregular criterion as it was during the five trials before criterion. Since OR rates increased dramatically after the irregular criterion for P&H's subjects, while omission rates decreased, this lends further support to the notion that omissions and OR errors are inversely related. Therefore, one explanation for the OR rate discrepancy lies in the fact that while P&H's subjects may have utilized guessing, and as a result exhibited higher OR rates, SDM does not guess and instead produces higher omission rates. Future work with SDM will address the issues of coding associations and guessing, hopefully resulting in simulations which more closely emulate human performance.

One final distinction should be made between the SDM simulations and P&H's experiment. According to David Palermo (personal communication, December 14, 1993), some of the P&H subjects experienced what he calls the "Aha phenomenon", that is, at some distinct point in the experiment, the subject would have the feeling that she or he hit upon the rule (in some cases, this feeling was verbalized). Palermo also suggests that there was a quantitative difference in these subjects' performances at this point. However, although the original data were available, there was no indication of which subjects experienced the "Aha phenomenon". Therefore, it was extremely difficult to substantiate the claim that performance significantly increased as a result of this phenomenon. Some of P&H's subjects did show an increase in performance on the regular pairs at some point during the experiment. For example, during trials 1 through 17, P&H subject #23 was correct on the regular items 22% of the time. During these trials, this subject never exceeded 50% correct, and on 13 of these early trials, regular performance was at or below 25%. In contrast, during trials 18 through 35, the same subject was correct on the regular items 74% of the time. During these later trials, subject #23 only fell below 50% correct on the regulars during one trial, and on 15 of these latter trials, regular performance was well above 60%. Based on Palermo's comments, subject #23 could have experienced the "Aha phenomenon" at trial 18, but this is only speculation. Nonetheless, since the possibility of the "Aha phenomenon" cannot be ruled out, SDM's ability to simulate this phenomenon must be addressed.

None of the SDM simulations of P&H showed any evidence of a dramatic increase in regular performance at a specific point during the simulation, in other words, the SDM simulations showed no evidence of the "Aha phenomenon". Therefore, should future research substantiate the existence of this phenomenon behaviorally, another direction for future research should be devoted to establishing SDM's ability to simulate it.

In summary, despite the differences outlined above, SDM was able to model a number of the most important characteristics of P&H's human data, adding further support to the view that SDM is at least a model of human memory worthy of additional study. The similarities between the human data and the simulations include:

1) In both situations the irregulars were learned earlier, and with overall greater accuracy, than the regulars.

2) In both P&H and the simulations, performance on the regulars was consistently below that of the irregulars, and there was a positive relationship between frequency of presentation of the irregular and rate of acquisition, so that as frequency of presentation increased, so did the acquisition rate.

3) Most importantly, both situations revealed shallow U-shaped learning accompanied by low levels of overregularization. There was also a positive association between frequency and correct performance and a negative

association between frequency and OR rates.

C. Friends and Enemies Experiment & Simulations

In 1979, Robert Glushko examined rules and exceptions in reading aloud. According to Glushko, most researchers at the time believed that readers use spelling-to-sound rules when they encounter novel words for which they do not possess stored pronunciations. This is similar to the picture presented for the acquisition of the past tense in which children use a rule to form the past tense of a novel verb. Just as there are exceptions to the regular past tense rule, there are also exceptions to the pronunciation rules. In his third experiment, Glushko (1979) identifies three different types of words, based on whether or not they follow the established pronunciation rules and whether their neighbors also follow the rule. In his analysis, neighbors are words which share the same "body" and only differ in their first consonant, for example, "haze" and "maze". Glushko's "regular consistent" words are pronounced regularly, and so are all their neighbors (e.g., "wade", whose neighbors include "jade" and "made"). "Regular inconsistent" words are words which are pronounced regularly, but some of their neighbors are not pronounced using the regular rule (e.g., "wave", whose neighbors include the regularly pronounced words "cave", "gave", "pave", "rave" and "save" and the irregularly pronounced word "have"). Finally, exceptions are those words that do not follow the regular pronunciation rules (e.g., "have"). Using a naming procedure, Glushko found that regular consistent words had the shortest pronunciation latency, followed by the regular inconsistent words and then the exceptions. He also found that the error rates were related to the naming latencies, with the highest error rate belonging to the exception words. Glushko also replicated this effect with non-words. Using regular consistent nonwords like "bink", which is derived from the regular consistent word "pink", and irregular inconsistent nonwords like "bint", which is derived from the regular inconsistent word "mint" (whose neighborhood includes the irregularly pronounced word "pint"), Glushko once again found that the naming latencies for inconsistent regulars was longer than that for consistent regulars. As further support of the theory that neighbors can effect the pronunciation of a word, Glushko found that on 8.7% of the trials, inconsistent regular nonwords were pronounced as exceptions (e.g., the word "bint" was pronounced like the exception "pint" rather than like the regular "mint"). Glushko used these results to develop an activation-synthesis model of pronunciation in which the presentation of a word or letter string first causes the activation of visually similar words and then results in the pronunciation of the word based on the synthesis of all activated words. As a result, words which activate entries which are all pronounced similarly are easier to pronounce than words which activate entries with various pronunciations.

Glushko's (1979) experiments had a dramatic effect on the traditional theories of pronunciation, just as Rumelhart and McClelland's (1986) simulations bred controversy in the past tense acquisition domain. His conclusions are also similar to those of Rumelhart and McClelland in a way which is very pertinent to this discussion. Glushko's activation-synthesis model proposed that a single process could be used to explain the pronunciations of both words and exceptions, just as Rumelhart and McClelland proposed that a single network could account for the formation of the past tense of both regular and irregular verbs. The traditional dual-route theory of pronunciation had held that both lexical (word based) and nonlexical (rule based) mechanisms were required to successfully pronounce words. Glushko, on the other hand, postulated that no pronunciation rules are required. Instead, pronunciation is the result of activation of all words within the input word's neighborhood, with the pronunciation of the input word resulting from a synthesis of these activated neighbors.

Glushko's (1979) work in the field of pronunciation was furthered by the research of Jared, McRae and Seidenberg (1990). In a series of experiments, they investigated the idea of neighborhoods in greater detail. They proposed that it is not only the consistency of a word's neighborhood, but more specifically, the relative frequencies of "friends" and "enemies" within the word's neighborhood, which determines the ease of pronunciation of the word. In the terminology of Jared et al., a "friend" is a member of the neighborhood of an input word which is pronounced like the input word. An "enemy", then, is a member of the input word's neighborhood which is not pronounced like the input word. Take the word "haste", for example. Its neighborhood consists of three friends, "waste", "paste" and "taste" and one enemy "caste" (in Glushko's terminology, "waste" would therefore be considered a regular inconsistent word and "caste" would be an exception). In their first three experiments, Jared et al., used a naming paradigm to examine the effects of friends and enemies on the pronunciation of words. They found that naming latencies depend on the relative frequencies of the friends and enemies in a word's neighborhood, but not on the relative number of friends and enemies. The results of their first experiment demonstrated that the presence of enemies slows down the naming process and increases error rates. The consistency effect (which is the difference in naming latencies between consistent and inconsistent words) was significant for words with high frequency enemies, but not for words with low frequency enemies. One unexpected result was the fact that words with low frequency enemies took longer to name than words with high frequency enemies. However, the words with high frequency enemies also had high frequency friends and the words with low frequency enemies also had low frequency friends, resulting in a possible confound. Therefore, in experiment 2 they crossed both frequency of friends (high vs. low) and frequency of enemies (high vs. low). Once again, the presence of enemies (whether of high or low frequency) hampered the naming process and caused more errors. Figure 30 shows the general pattern which emerged. Of all the words with enemies, words with low frequency friends and high frequency enemies took longest to name, followed by those with high frequency friends and high frequency enemies, low frequency friends and low frequency enemies, and high frequency friends and low frequency enemies, in that order. It therefore appears that the major contribution was supplied by the word's enemies, because as the frequency of enemies increased, naming latencies also increased. The effect of friends seems to have been secondary, with increasing frequency of friends facilitating naming. Experiment 3 investigated whether it was actually the frequency of the friends and enemies which had the effect, or if it was the number of friends and enemies. In this case, the results indicated that the size of the effect is in fact predicted by the frequency, and not the number, of friends and enemies. All these results clearly add further support to Glushko's (1979) activation-synthesis model which assumes that pronunciation does not necessarily

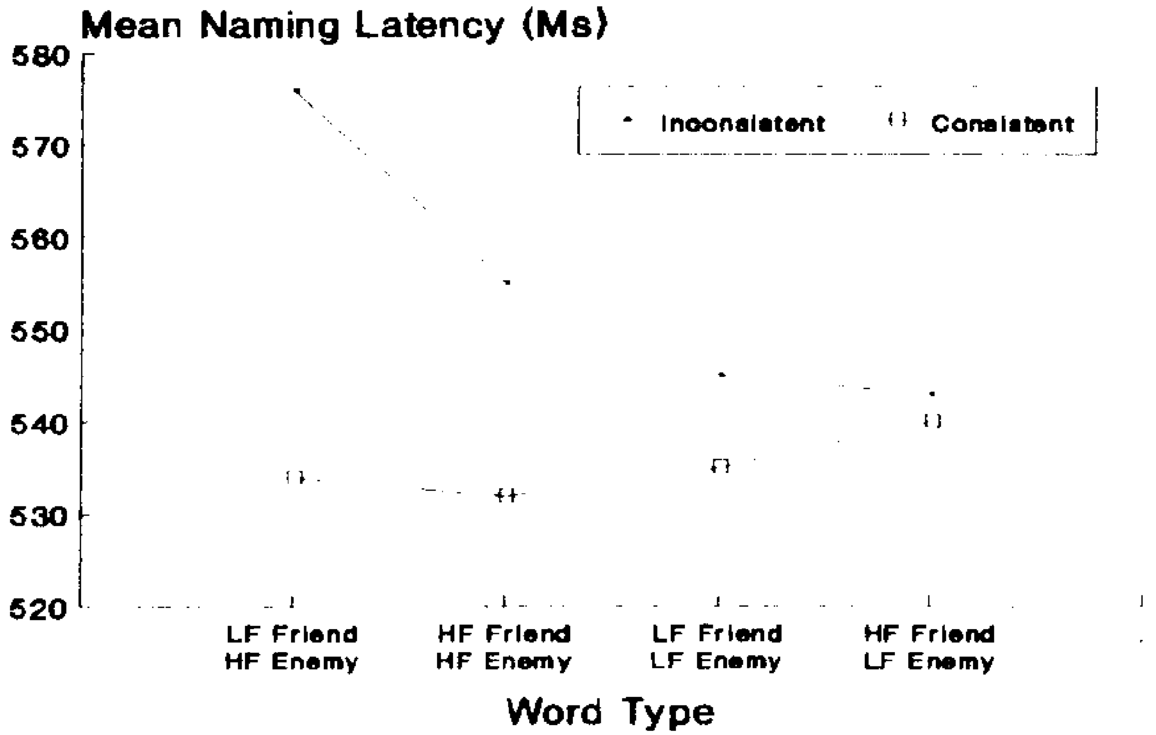


Figure 30. Jared et al.,(1990) Experiment 2: Mean naming latency in milliseconds for consistent and inconsistent words based on the frequency of friends and enemies in each word's neighborhood.

rely on the use of rules, but instead depends on a pronunciation which is the result of a synthesis of the pronunciation of items in a particular word's neighborhood.

At this point it should also be noted that all Jared et al., (1990) results were simulated using Seidenberg and McClelland's (1989) PDP model for word recognition and naming. For experiment 1, the phonological error scores generated by the model simulated the main effect quite closely, performing better on consistent words (without enemies) than on inconsistent words (with enemies). However, unlike Jared et al., results, the model did not exhibit the behavioral trend which indicated that words with high frequency enemies showed a larger consistency effect than words with low frequency enemies. The simulation of experiment 2, in which both frequency of friends and enemies were examined, indicated that just like their subjects in experiment 1, the model performed better on words without enemies than words with enemies. The largest effect exhibited by the model also followed the behavioral data, words with high frequency enemies and low frequency friends showed the most errors and the largest consistency effect[24]. However, once again, the model differed from the experimental results in a number of ways. First the graded

---

[24] The consistency effect is defined as the difference in naming latencies between inconsistent words (which have enemies) and their matched consistent words (which do not have enemies).

effects observed behaviorally were not exhibited by the model. As depicted in Figure 31, the model did not show any difference in consistency effects among words with high frequency friends and high frequency enemies, low frequency friends and low frequency enemies, and high frequency friends with low frequency enemies. Secondly, there was an indication that the model was more affected by the frequency of friends factor than the human subjects were.

1. Friends and Enemies Experiment

Since the experiments conducted by Glushko (1979) and Jared, McRae and Seidenberg (1990) indicate that naming, a process traditionally considered to require the use of both rules and exceptions, is influenced by the composition of a word's neighborhood, it seems reasonable to hypothesize that other rule based systems with exceptions might also be affected by neighborhood effects. The purpose of this experiment was to examine whether "friends" and "enemies" have an effect on the acquisition of a rule based system using a paired associate learning task similar to the one employed by Palermo and Howe (1970). In addition, since Palermo and Eberhart's (1968) and Palermo and Howe's (1970) results indicate that U-shaped learning and overregularization are not confined to the domain of language acquisition, this experiment is designed to lend further support to the claim that U-shaped learning and

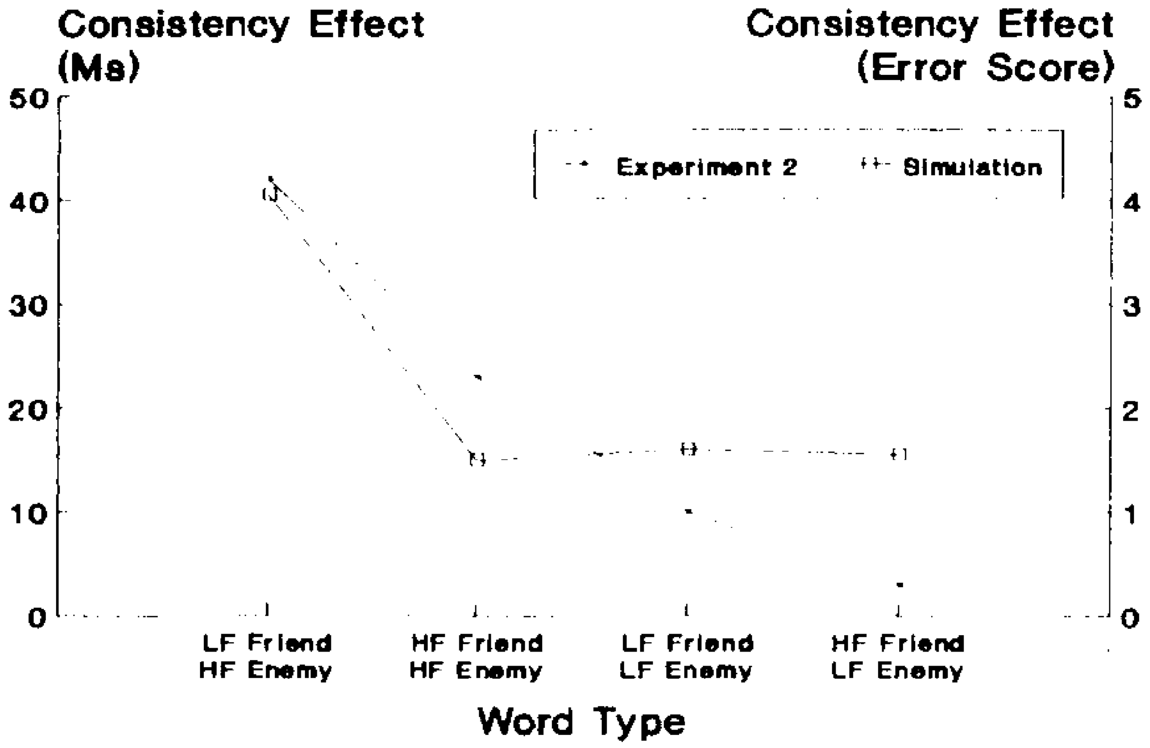


Figure 31. Jared et al.,(1990) Experiment 2 and simulation: The magnitude of the consistency effect in milliseconds (for Experiment 2) and error score (for the simulation) based on the frequency of friends and enemies in each word's neighborhood.

overregularization are general cognitive phenomena encountered in the acquisition of a system which could be described as having rules and exceptions.

a. Method and Procedure

Nineteen undergraduate and graduate students at Hunter College of CUNY served as subjects in the experiment; all were unpaid volunteers.

The stimuli in the experiment included 13 different pairs of three digit stimuli and responses. The pairs were designed to follow a rule based system very similar to R&M's rule of 78. The same restrictions were placed on the digits which could occupy each of the three positions in the stimuli and responses, and the same rule applied (refer back to section II.D.1.a for a detailed description). The pairs used in this experiment differed from those used in the rule of 78 in three important ways. First, three irregular pairs were used instead of one. Secondly, some pairs were not used. Finally, in order to study the frequency effects of friends and enemies, two pairs (one regular and one irregular) were presented three times per trial so that each subject saw a total of 17 pairs during each trial. Table 8 lists the 13 pairs used in the experiment.

Any two stimuli which share two digits in the same position are considered neighbors (e.g., "367" and "368"). Two stimuli are considered friends if they are neighbors and

Table 8

Friends & Enemies Experiment: Stimulus - Response Set

	STIMULUS	RESPONSE
REGULAR CONSISTENT PAIRS:	247	248
	258	257
	268	267
REGULAR INCONSISTENT PAIRS:	147	148
	167	168
	248	247
	267	268
	347	348
	358	357
	368	367
IRREGULAR PAIRS:	148	148
	357	357
	367	367

require the same last digit in their responses (e.g., "367"-> "367" and "357" -> "357"). Enemies are neighbors which require a different last digit in their responses (e.g., "367" -> "367" and "267" -> "268"). Therefore, in general, friends of irregular stimuli will also be irregular and enemies of irregulars will be regular. Since U-shaped learning and overregularization are primary concerns in this study, the neighborhoods of the irregular pairs are of most interest. One irregular (357-357) has a neighborhood which consists of a high frequency friend (the irregular pair 367-367, presented three times per trial) and a low frequency enemy (the regular pair 347-348). A second irregular pair (148-148) has a low frequency friend (the regular inconsistent pair 147-148) and a high frequency enemy (the regular inconsistent pair 248-247, presented three times per trial). The final irregular (367-367) has two friends (itself, since it is presented three times per trial and the irregular 357-357) and three low frequency regular inconsistent enemies (267-268, 167-268, 368-367). In addition, two of the irregular pairs (357-357 and 367-367) has one regular inconsistent neighbor which is neither its friend nor its enemy. Finally, in order to avoid a task so simple that no effects would be seen, three additional pairs, which are not in any of the irregulars' neighborhoods, were included. Table 9 outlines the neighborhoods of the three irregular pairs used in this

Table 9

Friends & Enemies Experiment: Irregular Neighborhoods

357 - 357: High Frequency Friend, Low Frequency Enemy

Friend: (irregular)	Low Frequency Enemy: (regular INC[25])
367 - 367	347-348
367 - 367	
367 - 367	Other Neighbors: (regular INC)
	358 -357

148 - 148: Low Frequency Friend, High Frequency Enemy

Friend: (regular INC)	Enemy: (regular INC)
147-148	248 - 247
	248 - 247
	248 - 247

367 - 367: Mixed Frequency Friends, Low Frequency Enemies

Friends: (irregular)	Enemies: (regular INC)
357 - 357	267 - 268
367 - 367	167 - 168
367 - 367	347 - 348

Other Neighbors: (regular INC)
368 - 367

Regular CON[26] Pairs: Not Neighbors of the 3 Irregulars

247 - 248
258 - 257
268 - 267

---

[25] INC = inconsistent

[26] CON = consistent

experiment.

Each subject in the experiment was run individually on a microcomputer. Subjects began by reading the instructions for the experiment on the computer screen. The text of the instructions may be found in Appendix B.

Subjects were permitted as much time as they needed to read and understand the instructions. Once all questions were answered by the experimenter, the subject was left alone for the remainder of the experiment. During the study period, each stimulus was displayed alone for 1 second, then the corresponding correct response was displayed and the pair remained on the screen for 3 seconds. There was a 2 second interval between the presentation of each successive pair. Each subject received a unique randomized presentation order of the 17 total pairs during the study period. Upon completion of Part I, "End of study period" was displayed on the screen for 2 seconds, followed by a 2 second presentation of "Get ready for the learning phase". During each trial of the learning phase, one of the three digit stimuli was displayed on the screen for up to 7 seconds. The subject responded using the number pad on the keyboard. Once a response was made, or after 7 seconds if no response was made, the stimulus number, the correct response and the subject's response were displayed on the screen for 1.5 seconds. There was then a 2 second interval before the next stimulus number was presented. The presentation order of the stimuli was randomized on each trial for each subject in the learning phase. After all 17 stimuli were presented, "End of trial" was displayed for 2 seconds and then the next learning trial began. The subject's overall percent correct on each learning trial was calculated by the computer and the experiment terminated once the subject had participated in at least 20 learning trials and had at least one trial at 100% correct on all pairs, or, if this criterion was not reached, after 30 trials.

Upon completion of the experiment, some subjects were questioned about what they thought they had learned in the experiment. Reaction times were recorded for eight of the nineteen subjects[27]. In addition, two of the 19 subjects were run as protocol subjects. These two subjects were given as much time as they wanted to respond, instead of the 7 second limit placed on all other subjects, and were asked to "think out loud" about how they decided on each of their responses. Their thoughts were recorded in order to try to ascertain exactly how strategies developed and changed

---

[27] Reaction times were not recorded for all subjects since originally they were not considered critical to the analysis of the experiment. However, at the suggestion of a reader, the experimental program was changed to collect RTs for the last 8 subjects.

throughout the course of the experiment[28].

b. Results

Twelve subjects were run for twenty trials each, since they reached a criterion of one errorless trial by this point. Four subjects were run for twenty five trials[29] and three subjects were run for the maximum thirty trials. Two of the subjects who were run for the maximum number of trials never reached 100% correct on all pairs[30].

In some of the analyses performed for this experiment, only the last ten trials of each subject's experiment were used. This decision was made in order to be able to make more meaningful comparisons to Jared et al., (1990) results. Both Glushko (1979) and Jared et al., studied naming latencies for word pronunciation. The assumption is that all their subjects had previous experience with the words used, and therefore learning was not a component in the

---

[28] During the course of running subjects for this experiment, it was unclear exactly what strategies the subjects were using to learn the stimulus-response set. Preliminary analyses did not clearly indicate whether the subjects were discerning the "rule of 78" or the concepts of friends and enemies. Therefore, at the suggestion of a reader, two subjects were run as protocol subjects in order to try to get a clearer picture of the acquisition process.

[29] Two of the subjects who ran for 25 trials did not reach a criterion of 100% correct on all pairs. One reached 88.24% correct (15/17 correct) and the other, 82.35% correct (14/17 correct). The experiment had to be terminated at twenty five trials for these two subjects because of time considerations.

[30] The maximum percent correct for all pairs achieved by these two subjects was 88.24% correct (15/17 correct) and 82.35% correct (14/17 correct) respectively.

experiment. However, in this experiment, subjects were presented with a previously unencountered set of stimulus-response pairs to learn[31]. Each subject's overall percent correct per trial was examined to determine acquisition patterns and it was decided that the last ten trials were most indicative of their post- learning performance. For all nineteen subjects the overall mean percent correct on the last ten trials was 90% (range = 69.4% - 99.4%).

Mean percent errors during the last ten trials for the irregular pair (148-148) with a low frequency friend and high frequency enemies is compared to that of the irregular pair (357-357) with high frequency friends and a low frequency enemy in Figure 32. The third irregular (367-367) was left out of this analysis since it has both high frequency and low frequency friends and would therefore confound the effect of main concern here. The mean percent error score was 17.89% for the high frequency friend-low frequency enemy (HFF-LFE) pair, and 25.79% for the low frequency friend-high frequency enemy (LFF-HFE) pair. Nine subjects showed fewer errors for the HFF-LFE pair, two had fewer errors for the LFF-HFE pair, and eight subjects showed no difference in the mean % errors between the two conditions. Therefore the difference was only marginal by a

---

[31] The subjects chosen for this experiment had no previous knowledge of Rumelhart and McClelland's (1986) "rule of 78" in order to make sure that all subjects would actually have to learn the stimulus-response set.

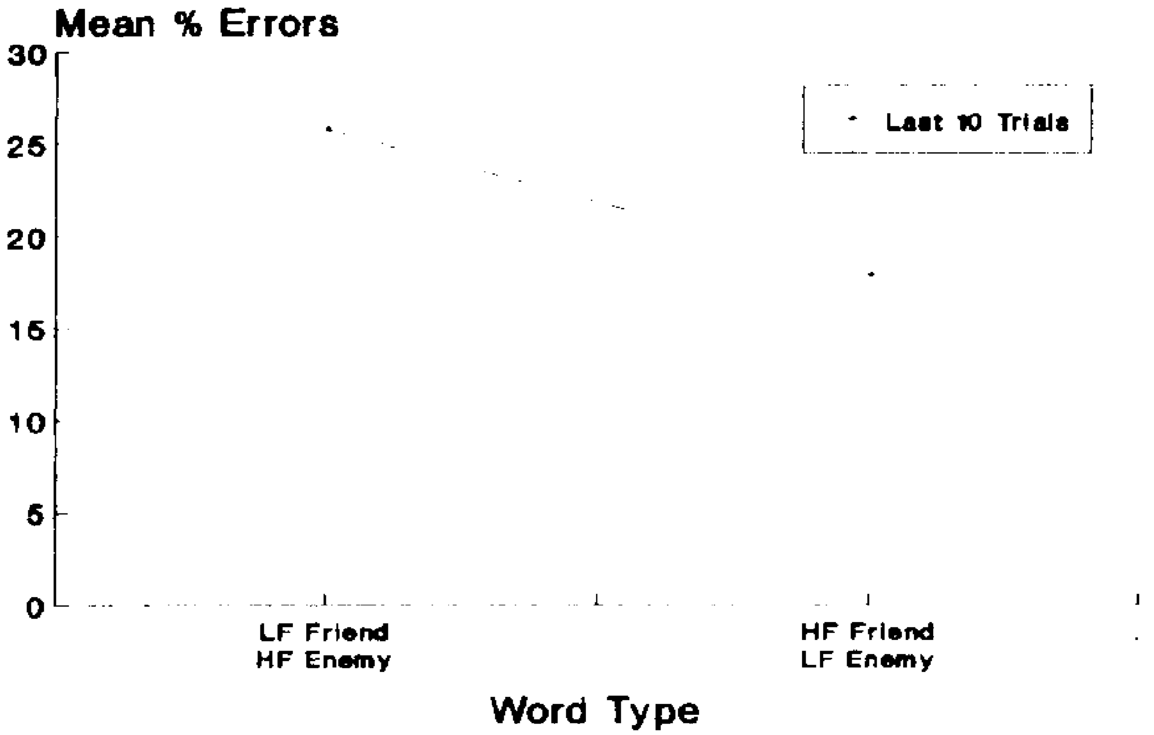


Figure 32. Friends and Enemies (F&E) experiment: Mean percent errors for irregular pairs during the last ten trials based on the frequency of friends and enemies in each pair's neighborhood.

dependent t-test (performed on the arcsine transformations of the proportions) $t(18) = 1.923$, $p<.10$, but was significant by a sign test, $p<.05$[32].

Frequency effects were also examined. In this analysis (and the following consistency analysis), only 17 subjects were included since the first two subjects run in this experiment were run without the three regular consistent pairs (see Table 9). Figure 33 illustrates the effects of frequency for regular and irregular pairs over all trials. A two way ANOVA with repeated measures was performed on the arcsine transformations of the proportions and indicated a significant frequency effect, $F(16,1)= 65.947$, $p<.001$, a significant regularity effect, $F(16,1)= 11.185$, $p<.005$, and a significant interaction, $F(16,1)= 23.647$ $p<.001$. Planned comparisons (also performed on the arcsine transformations of the error proportions) revealed that high frequency (HF) irregulars (mean = 15% errors) showed significantly fewer errors than low frequency (LF) irregulars (mean= 37.75% errors), $t(16)= 6.601$, $p<.001$. However, the difference between HF regulars (mean = 14.05% errors) and LF regulars (mean = 16.08% errors) was only marginal, $t(16)= 2.103$, $p<.10$.

Consistency effects over all trials are illustrated in Figure 34. A consistent pair is defined as a regular pair

---

[32] There was no significant difference in reaction times between the HFF-LFE and LFF-HFE pairs for the 8 subjects for whom RTs were recorded, $t(7)= 0.88$, $p>.10$.

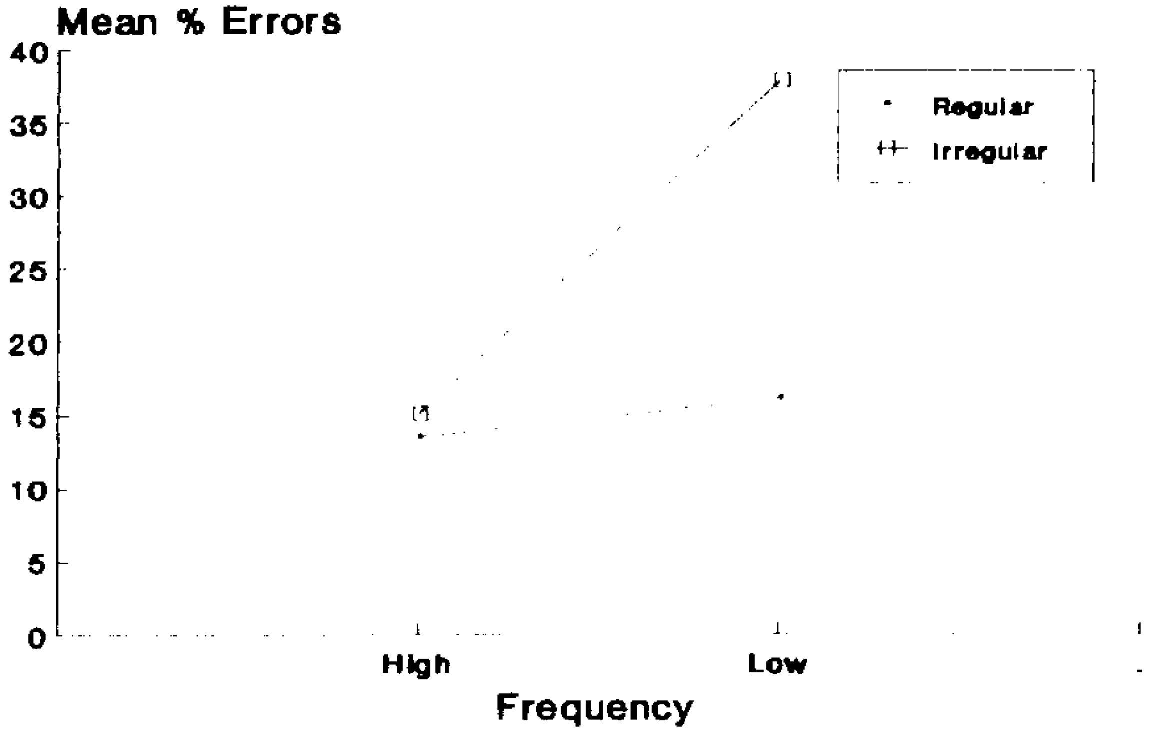


Figure 33. Friends and Enemies (F&E) experiment: Frequency effect. Mean percent errors for regular and irregular pairs based on the frequency of presentation of each pair.

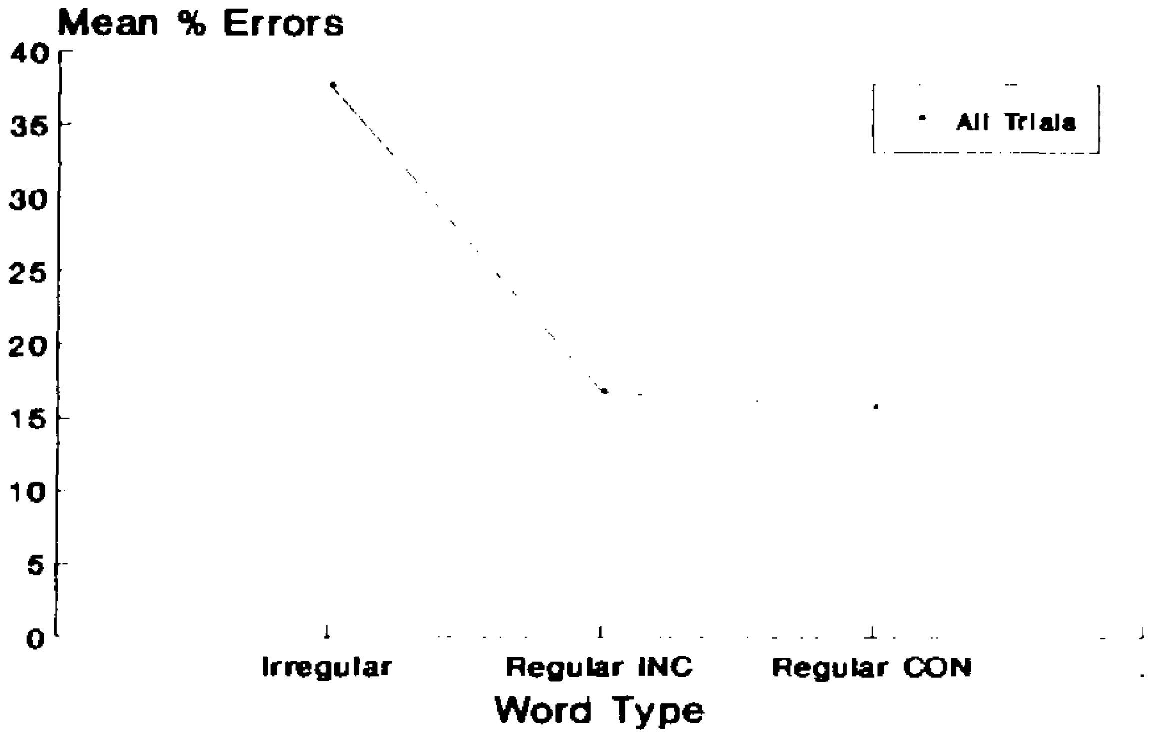


Figure 34. Friends and Enemies (F&E) experiment: Consistency effect. Mean percent errors over all trials for irregular, regular inconsistent (INC) and regular consistent (CON) pairs.

which only has regular pairs in its neighborhood. Therefore, a regular consistent pair cannot be a neighbor of any of the irregular pairs. Table 9 identifies each regular pair as consistent (CON) or inconsistent (INC). In this analysis, the HF irregular (367-367) and HF regular (248-247) pairs were excluded in order to avoid confounding the consistency effect with frequency effects. A repeated measures ANOVA with the three word types treated as levels (performed on the arcsine transformations of the error scores) indicated a significant consistency effect over all trials, $F(2,16)= 17.791$, $p<.001$. Planned comparisons revealed that the irregular pairs had significantly more errors than both the regular INC, $t(16)= 4.808$, $p<.001$, and regular CON pairs, $t(16)= 4.199$, $p<.001$. Error rates for regular INC and regular CON pairs were not significantly different.

The next matter of interest is whether or not these subjects showed U-shaped learning and overregularization while learning this stimulus - response set. During the acquisition process it took subjects an average of 8.3 trials to reach criterion on the irregular pairs, 8.6 trials for the regulars, and 11.6 trials for all pairs[33]. Figure 35 shows the mean percent correct for the irregulars and regulars, and the mean percent of OR errors, starting at the

---

[33] As in P&H (1970), criterion is considered to be the first trial at 100% correct.

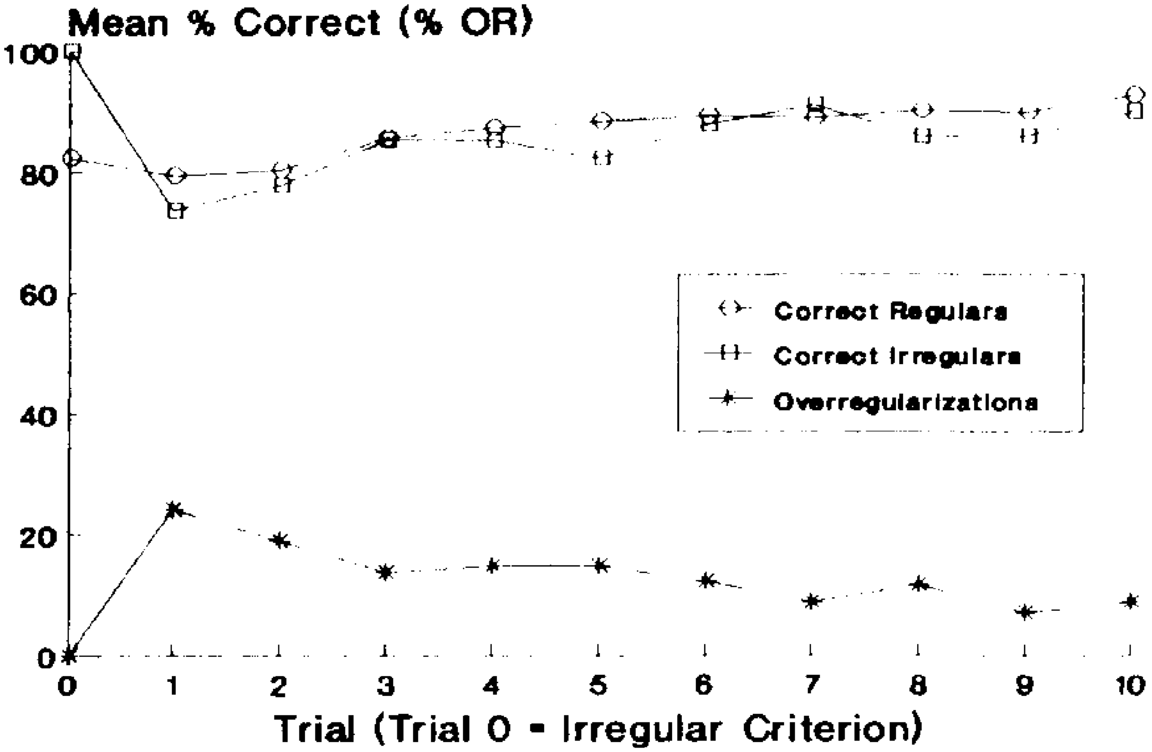


Figure 35. Friends and Enemies (F&E) experiment: Mean percent correct/trial for regular and irregular pairs and mean percent OR errors/trial for irregular pairs, beginning with the trial during which each subject first reached 100% correct on the irregular pairs (trial 0).

point at which each subject reached criterion on the irregulars (trial 0). While the regulars show consistent improvement during this period (an increase in performance of 10.2% over the 10 trials following the irregular criterion), the irregulars exhibit U-shaped learning (a decrease in performance of 26% at the lowest point) from trials 0 through 3. This U is accompanied by a concurrent increase in the OR rate (an increase of 24% at trial 1). Overregularization errors accounted for 86% of all errors before criterion was reached on the irregulars, and 79% of all errors after the same criterion. The right hand arm of the U is evident from trials 3 through 10, where the irregulars showed a fairly consistent increase (4.7%) in performance, until, by trial 10, subjects were performing at 90% correct for all irregulars.

Finally, the overall acquisition process for all subjects is illustrated in Figure 36. Over all trials, the regulars showed consistently superior performance when compared to the irregulars. As performance improved on the irregulars, there was a simultaneous decrease in the overregularization rate. Over all trials, OR errors accounted for 88% of all irregular errors.

2. Friends and Enemies Simulations

a. Method

In contrast to the simulations of the "rule of 78" and

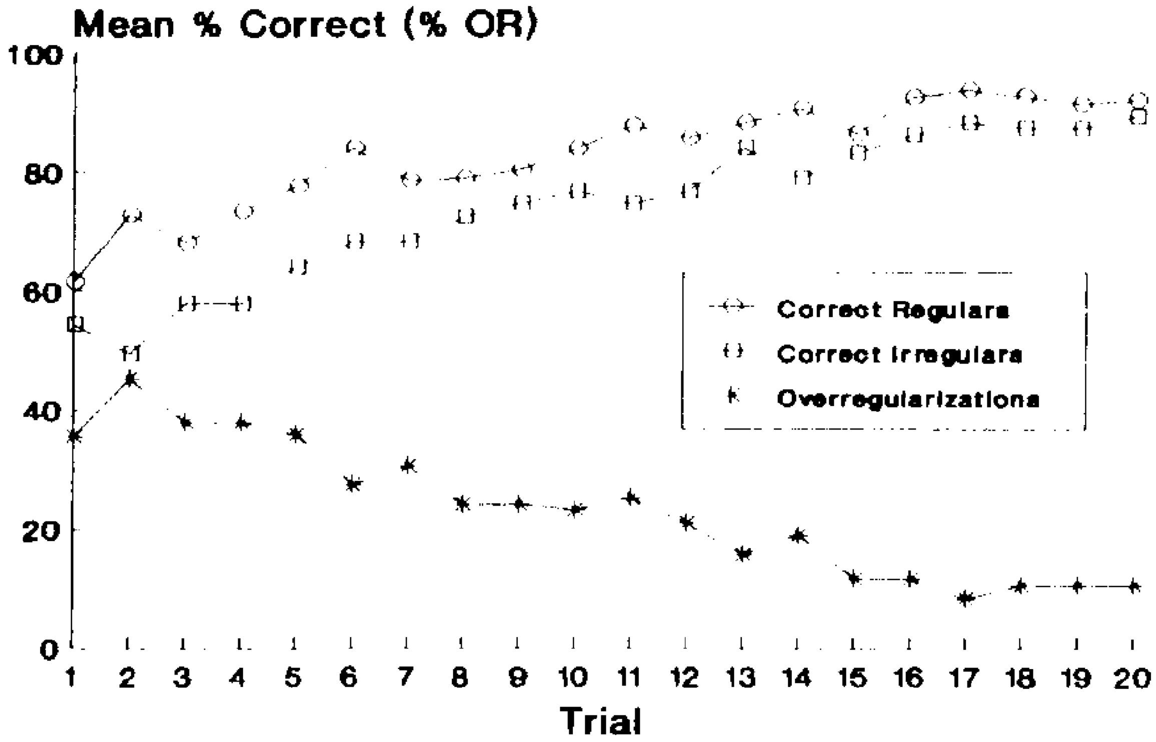


Figure 36. Friends and Enemies (F&E) experiment: Mean percent correct/trial for regular and irregular pairs and mean percent OR errors/trial for irregular pairs.

Palermo and Howe (1970), in the simulations of the friends and enemies experiment (hereafter referred to as F&E), the responses coded in the simulations did not exactly match the responses required of the human subjects. Instead of coding the whole response, only the last digit of the response (7 or 8) was coded[34]. The data vector representing the last digit of the correct response was stored at the vector address of the stimulus. Sixteen-bit vectors were used to represent both the stimulus and the single digit response. The last digit of the responses was coded with non-overlapping bits in the data vectors. Table 10A shows the 16-bit vectors for each of the responses. Each response (7 or 8) was coded by eight unique bits in the vector so that the Hamming distance between the two responses was 16 bits. Prior to the addition of noise, each digit in the stimulus address vector was coded by two unique bits so that the hamming distance between any two stimulus digits was 4 bits. Table 10B shows the position of the two bits turned "ON" (indicated by a "1") for each digit of the stimuli. Table 10C lists the actual address bit-vectors (prior to the addition of noise) for each of the 13 stimuli used in these

---

[34] Many preliminary simulations, which varied all the parameters, revealed that if the whole response was coded in this implementation of SDM, the simulations did not show the high overregularization rates indicated in the human results. This limitation of the program, and justification for the decision to only code the last digit of the responses in the simulations, will be addressed in greater detail in the discussion which follows.

Table 10

SDM Simulations of the Friends & Enemies Experiment

A. 16-Bit Vectors used to represent RESPONSES

RESPONSE	BIT-VECTOR
7	1111000011110000
8	0000111100001111

B. "ON" Bit Positions of each Stimulus DIGIT in the Address Vectors

STIMULUS DIGIT	ADDRESS VECTOR
1	1.......1.......
2	.1.......1......
3	..1.......1.....
4	...1.......1....
5	....1.......1...
6	.....1.......1..
7	......1.......1.
8	.......1.......1

C. 16-Bit Address Vectors used to represent STIMULI

STIMULUS	ADDRESS BIT-VECTOR
147	1001001010010010
148	1001000110010001
167	1000011010000110
247	0101001001010010
248	0101000101010001
258	0100100101001001
267	0100011001000110
268	0100010101000101
347	0011001000110010
357	0010101000101010
358	0010100100101001
367	0010011000100110
368	0010010100100101

simulations. The mean Hamming distance between any two stimuli was 7.9 bits with a range of 4 - 12 bits. Figure 37 depicts two examples of the graphical representation of the stimulus and response vectors used in the SDM simulations of the F&E experiment.

Five thousand "hard" memory locations were used in the simulations of the F&E experiment. During training, responses were stored at all addresses within a radius of 2 from the designated stimulus address. As in the two sets of previously reported simulations ("rule of 78" and P&H), a radius of 2 was chosen to include 1/1000th of the total memory space. Ten simulations were run. Each simulation consisted of 70 trials. The only difference between the ten simulations was in the seed used to initialize the memory space. During training, one irregular pair ("367-7") and one regular pair ("248-7") were presented three times per trial. The remaining pairs were presented one time per trial during training. Therefore, during training, the program was presented with 17 pairs per trial. During testing, the data contained in each of the 17 stimulus addresses[35] were read once per trial. During both training and testing the noise level added to each stimulus address vector was initially set to 0.30 and was reduced by 0.01

---

[35] During testing the data was read from the stimulus addresses using the same frequency conditions as during training. That is, data was read from addresses "367" and "248" three times per trial, and the data from all other stimulus addresses was read once per trial.

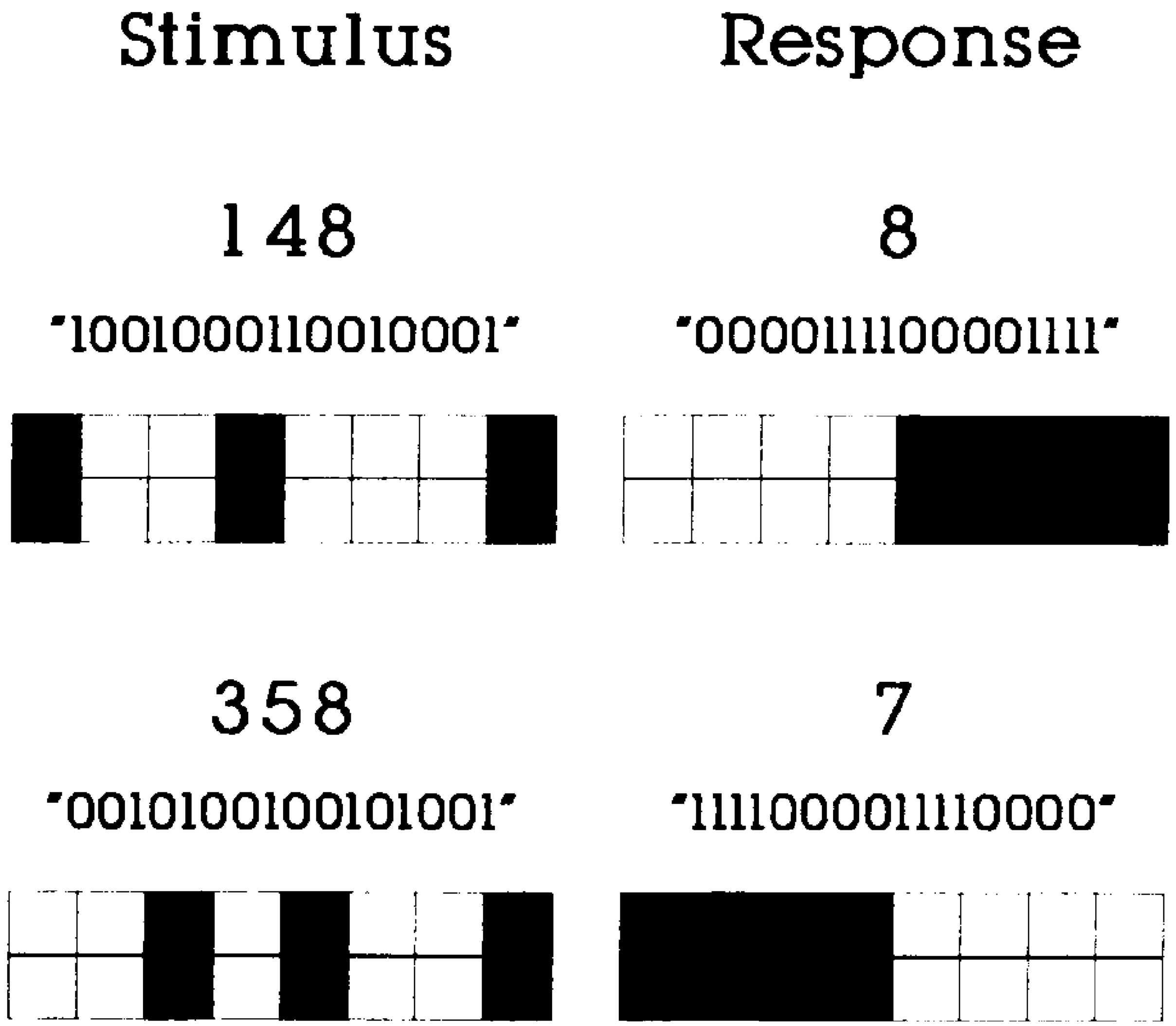


Figure 37. Graphical representation of the stimulus - response pairs used by the program in the SDM simulations of the F&E experiment. Two rows of 8 boxes were used to represent the 16 bit vectors. For the stimuli, each column of boxes represents the two bits used to represent each digit in the stimulus. Each response was represented by eight bits. Each box represents one bit. A black box represents a "1" in the bit vector and indicates that the bit is turned on. A white box represents a "0" in the vector and indicates that the bit is off.

every two trials until, during training, the level reached 0.02, and during testing, the noise level reached 0. Therefore, during training, trials 57 through 70 had a noise level of 0.02, and during testing, trials 61 through 70 had no noise added to the stimulus address vectors. Table 11 summarizes the parameters used in the 10 simulations of the F&E experiment.

b. Results

Figure 38 compares the mean percent errors over all trials for the irregular pair (148-8) with no friends and a high frequency enemy (LFF-HFE) to that of the irregular pair (357-7) with a high frequency friend and no enemies (HFF-LFE). A t-test, performed on the arcsine transformations of the error scores, revealed that over all trials, the LFF-HFE pair showed significantly more errors than the HFF-LFE pair, $t(9) = 11.982$, $p<.001$. The mean percent error score was 31.32% for the LFF-HFE pair and 13.71% for the HFF-LFE pair. Over all trials, all ten simulations showed fewer errors for the HFF-LFE pair than for the LFF-HFE pair. In order to make a comparison between these simulations, the F&E experiment and Jared et al., (1990), Figure 38 also compares the mean percent errors for the LFF-HFE and HFF-LFE pairs over trials 22-51. Subjects never reached consistently perfect performance in the F&E experiment, while these simulations all asymptote at 100% correct. Therefore, the

Table 11

Parameters for the SDM Simulations of the Friends & Enemies Experiment

Vector length	16 bits
Number of hard memory locations	5000
Seed	randomly generated by the program
Number of trials per simulation	70
Training:	
Stimulus - Response Presentations (storing response at stimulus address)	1 irregular presented 3x/trial 1 regular presented 3x/trial 2 irregulars presented 1x/trial 9 regulars presented 1x/trial
Radius	2
Noise added to stimulus address vectors	Trial 1: Noise = 0.30, and is reduced by 0.01 after every 2 trials until Noise = 0.02(trials 57-70)
Testing:	
Response Readings (retrieving response from stimulus address)	Same as training
Radius	Same as training
Noise added to stimulus address vectors	Trial 1: Noise = 0.30, and is reduced by 0.01 after every 2 trials until Noise = 0.00 (trials 61-70)

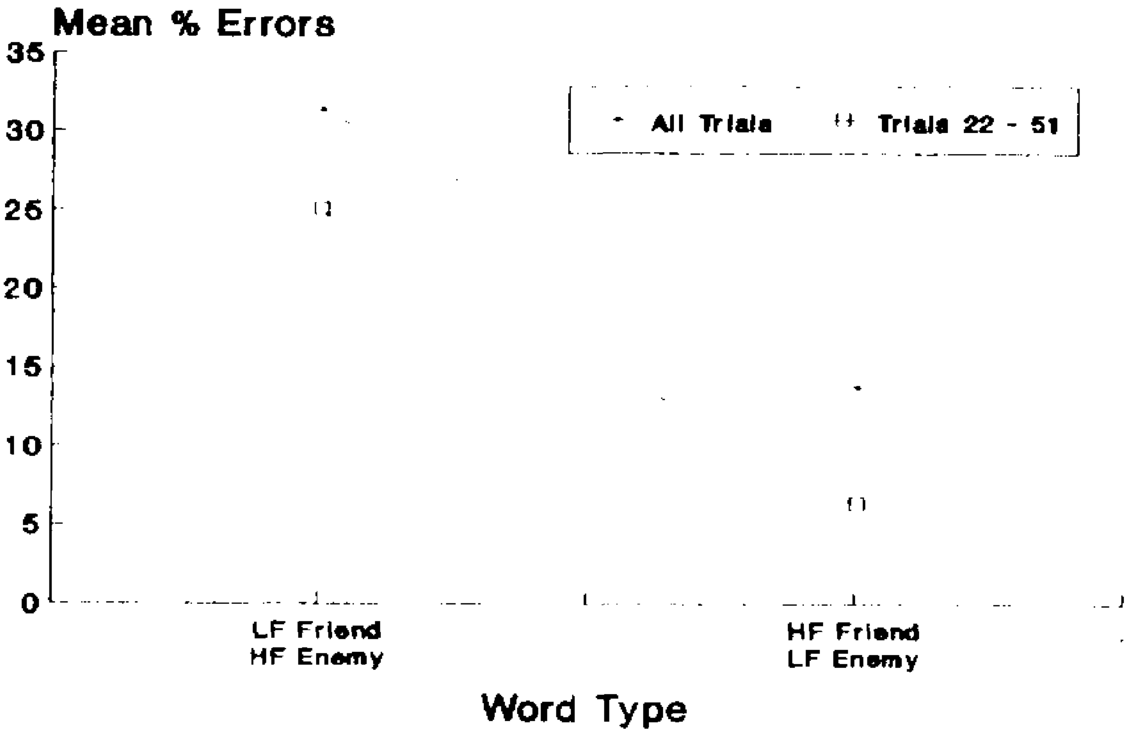


Figure 38. SDM simulations of the F&E experiment: Mean percent errors for irregular pairs over all trials and for trials 22-51, based on the frequency of friends and enemies in each pair's neighborhood.

last trials of the simulations cannot be directly compared to the last trials of the F&E subjects. Since the overall percent correct for the last 10 trials for the F&E subjects was 90%, and since the simulations were run for approximately three times as many trials as the F&E subjects[36], trials 22-51 were chosen for comparison. These simulation trials had a mean overall percent correct of 89.97%, with a range of 87.4% - 92.5%. As for all trials in the simulations, the LFF-HFE pair showed significantly more errors than the HFF-LFE pair, t(9)= 7.222, p<.001 during trials 22-51[37]. The mean percent errors was 6.33% for the HFF-LFE pair and 25% for the LFE-HFE pair. In all simulations, during trials 22-51, the HFF-LFE pair showed fewer errors than the LFE-HFE pair.

Figure 39 indicates the effects of frequency for regular and irregular pairs over all trials in the simulations. A two way ANOVA with repeated measures (performed on the arcsine transformations of the proportions) indicated a significant frequency effect, F(9,1)= 134.769, p<.001 and a significant regularity effect, F(9,1)= 12.682, p<.01, but no significant interaction. Planned comparisons (again, using arcsine transformations)

---

[36] The simulations were run for 70 trials each, while the F&E subjects were run for an average of 22.6 trials (range = 20-30).

[37] As for the analysis over all trials, the t-test for trials 22-51 used the arcsine transformations of the error scores.

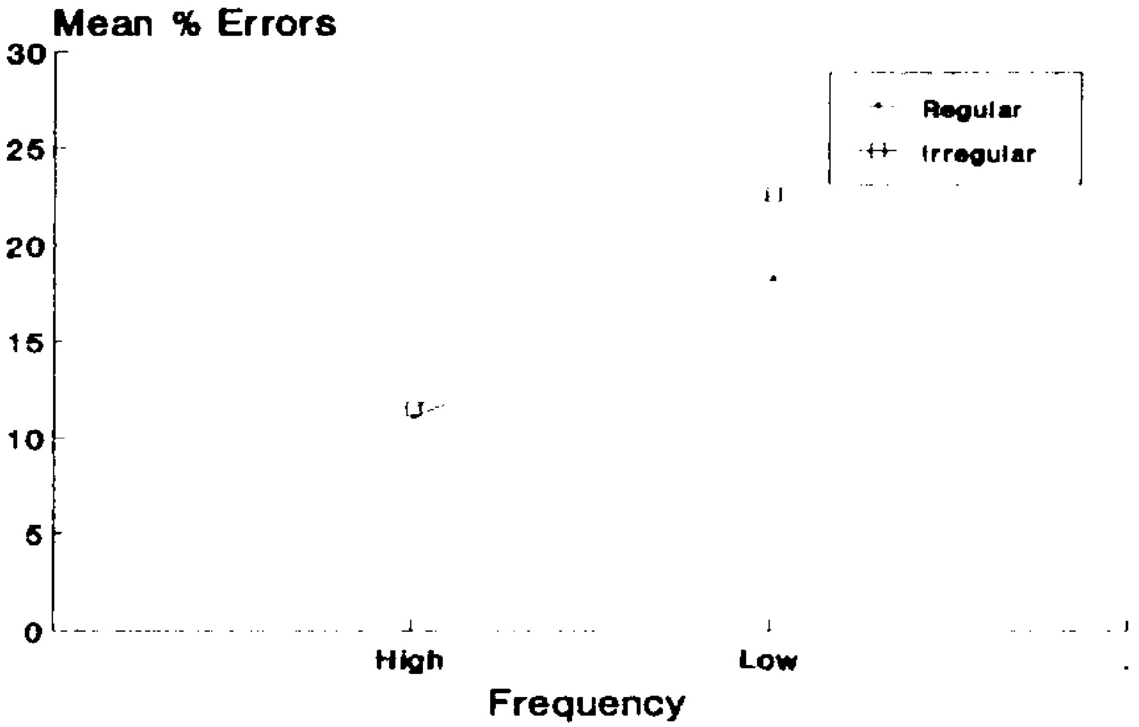


Figure 39. SDM simulations of the F&E experiment: Frequency effect. Mean percent errors for regular and irregular pairs based on the frequency of presentation of each pair.

revealed that high frequency irregulars (mean = 11.47% errors) showed significantly fewer errors than low frequency irregulars (mean= 22.5% errors), t(9)= 7.020, p<.001. Similarly, high frequency regular pairs (mean= 11.04% errors) showed significantly fewer errors than low frequency regular pairs (mean = 18.15% errors), t(9)= 7.727, p<.001.

Consistency effects over all trials are illustrated in Figure 40. As in the previously reported analysis of the F&E experiment, the HF irregular (367-7) and the HF regular (248-8) were excluded from this analysis, in order to avoid confounding consistency effects with frequency effects. A one way repeated measures ANOVA with three levels (word type) was performed on the arcsine transformations of the error scores and indicated a significant consistency effect over all trials, F(2,9)= 11.358, p<.005. Planned comparisons, utilizing arcsine transformations, revealed that regular CON pairs had significantly fewer errors than both irregular, t(9)= 4.094, p<.01 and regular INC pairs, t(9)= 5.994, p<.001. The difference between irregular and regular INC pairs only approached significance, t(9)= 1.20, p>.10.

During acquisition, the simulations took an average of 12.5 trials to reach a criterion of one errorless trial for the irregular pairs, 27.8 trials for the regulars, and 32.1 trials for all pairs. Figure 41 illustrates that U-shaped learning, accompanied by overregularization, was evident in

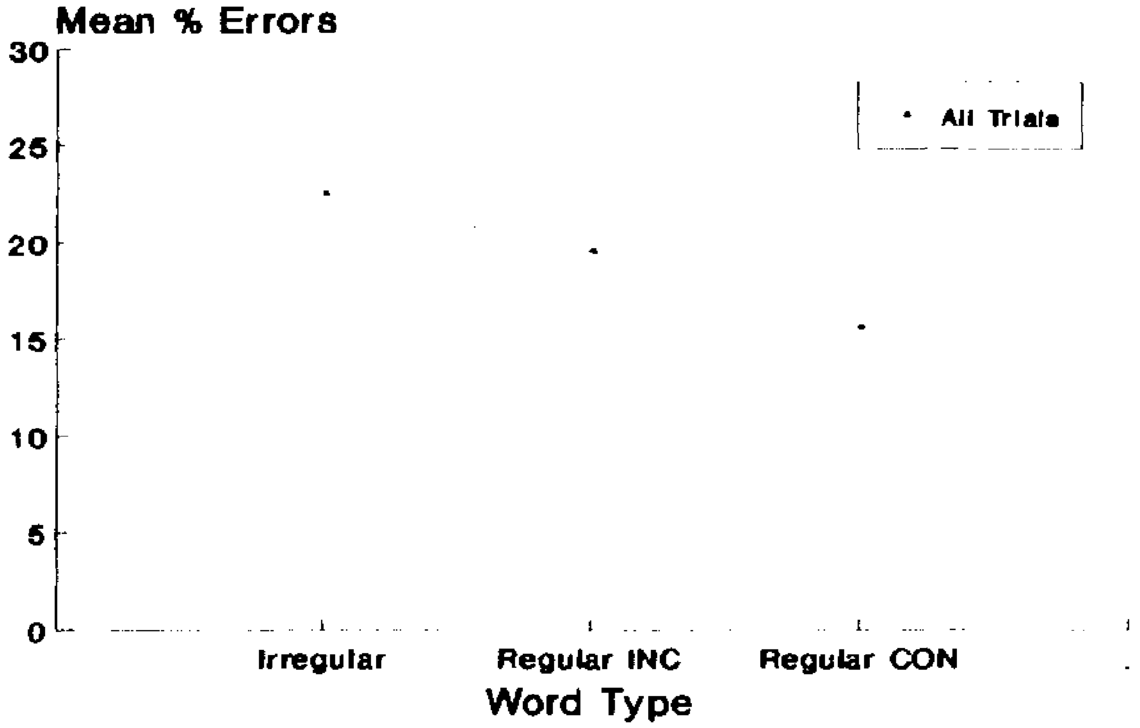


Figure 40. SDM simulations of the F&E experiment: Consistency effect. Mean percent errors over all trials for irregular, regular inconsistent (INC) and regular consistent (CON) pairs.

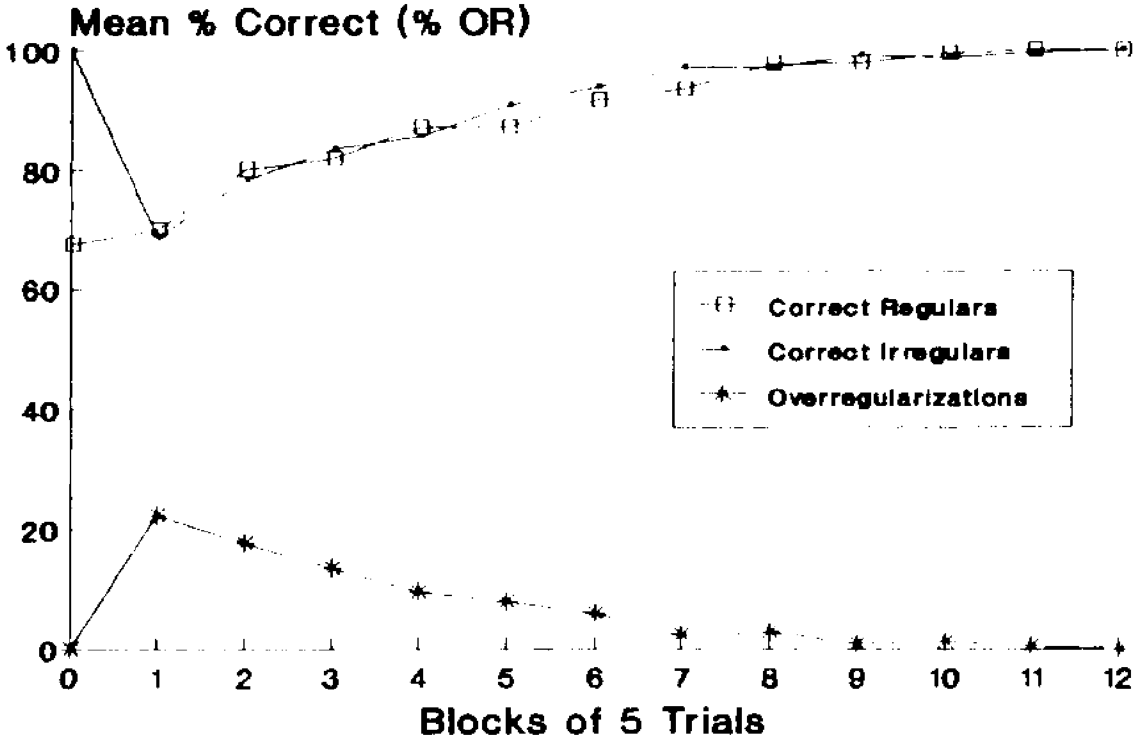


Figure 41. SDM simulations of the F&E experiment: Mean percent correct/block of 5 trials for regular and irregular pairs and mean percent OR errors/block of 5 trials for irregular pairs, beginning with the trial during which each simulation first reached 100% correct on the irregular pairs (trial 0). Trial 0 = 1 trial.

these simulations. As in all previous illustrations of U-shaped learning, trial 0 is the point at which each simulation first reached 100% on the irregular pairs. Performance for the regulars consistently improved during the trials following irregular criterion (a gradual increase in performance of 32% over 60 trials). In contrast, for the five trials following criterion the irregulars showed a decrease in performance of 40% and then gradually improved during the next 55 trials. This U was accompanied by a concurrent increase in the OR rate (an increase of 32% for the five trials following criterion). Overregularization errors accounted for 47.3% of all irregular errors before criterion and 81.3% of all irregular errors after criterion. Blocks 2 through 12 represent the right hand arm of the U, and irregular performance eventually returned to 100% during trials 52 through 64 following criterion (blocks 10-12).

Finally, overall acquisition for all 10 simulations is illustrated in Figure 42. Over all trials, performance for the irregulars and regulars is virtually indistinguishable. As performance improved for the irregulars, there was a simultaneous decrease in the number of overregularization errors. Over all trials, OR errors accounted for 63% of all irregular errors.

3. Discussion

Before discussing how well the SDM simulations model

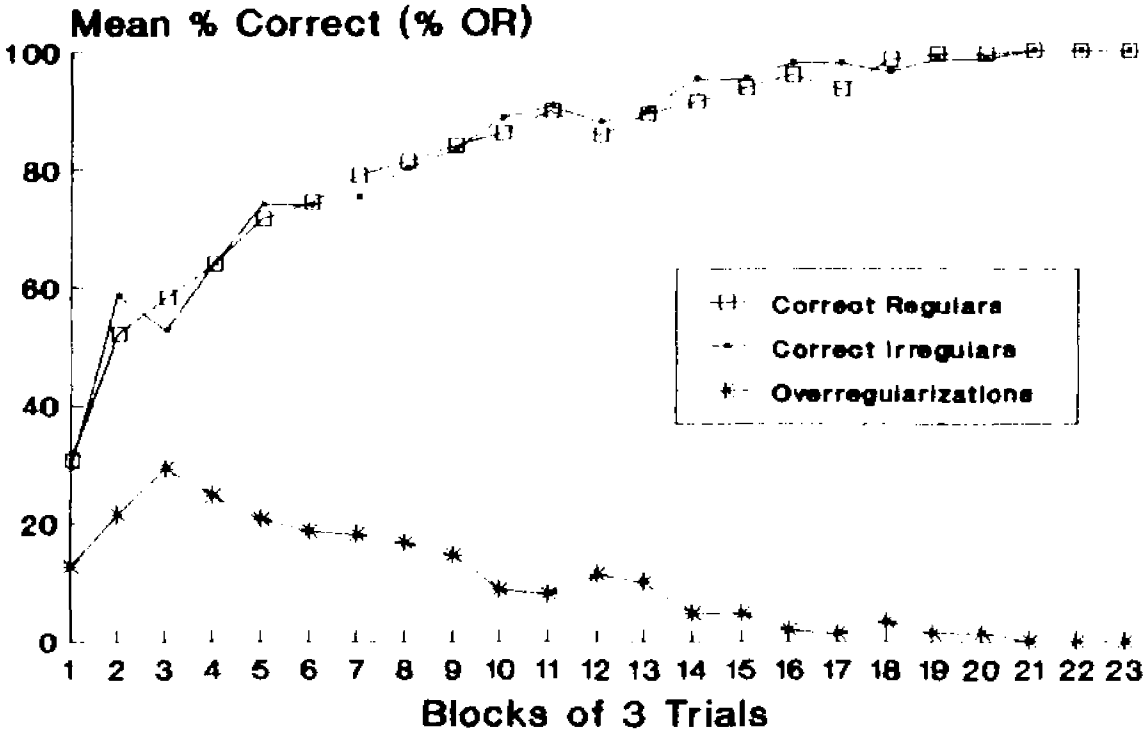


Figure 42. SDM simulations of the F&E experiment: Mean percent correct/block of 3 trials for regular and irregular pairs and mean percent OR errors/block of 3 trials for irregular pairs.

the human behavior represented by the F&E experiment, some mention must be made of the decision not to represent, in the simulations, the complete three digit response required of the F&E subjects. Initially, numerous attempts were made to code the entire response in the simulations. Unfortunately, these attempts were unsuccessful in one important way. Over all trials, OR errors accounted for 88% of all irregular errors made by the F&E subjects. In the simulations, when representations of the three digit responses were stored in addresses representing the three digit stimuli, little if any overregularization was seen. All parameters were varied, including the number of memory locations, the amount of noise added to the stimulus addresses, the vector length, the seed used to initialize the memory space, the storage and testing radii, and even the number of presentations per trial of the HF regular and HF irregular pairs. However, varying each of these parameters, either singly or in combinations, did little to increase OR rates in the simulations. Upon detailed scrutiny of the memory spaces resulting from numerous simulation runs, it was observed that if the irregular addresses were probed using a radius which encompassed 100% of the memory locations, a sort of overregularization became evident. For example, in a simulation which used 16 bit vectors and 5000 memory locations, if the stimulus address vector for 148 was read using a radius of 14 (which addresses all 5000 memory locations), the digit 7 was retrieved, while probing with a radius of 2 (which encompasses the standard 1/1000th of the memory space) produced a response of 148. Therefore, it appeared as if the model was overregularizing in a gross sense. In addition, both protocol subjects, and all 8 of the remaining 17 subjects who were questioned about what they learned following the completion of the experiment, indicated that very early in the experiment (as early as trial 1) they realized that in each pair, the first two digits of the stimulus and the first two digits of the response always matched. They also said that it was clear that each response could only (possibly) be different from its corresponding stimulus in the last digit position. These two factors led to the decision to represent only the last digit of the response in the simulations.

Coding only the last digit (7 or 8) of the response resulted in simulations which showed significant levels of OR, thereby providing a way of more closely modelling the human behavior. However, why this solution worked remains to be addressed. The most likely explanation lies in frequency effects. As previously mentioned, in the simulations of the "rule of 78", little OR was seen if the frequency of the irregular was greater than two presentations per trial. In addition, in Palermo and Howe (1970), and in the simulations of P&H, as the frequency of the irregular increased, the OR rate also decreased. In both these cases, the proportion of low frequency pairs to high frequency pairs was much higher than in these simulations of F&E. For the simulations of the "rule of 78", one irregular was presented twice per trial, while 17 regulars were each presented once per trial (in other words, there were 17 regular types and only 1 irregular type). Therefore, on each trial, the model was presented with 19 pairs, 17 of which were regular with a token frequency of 1, and one of which was an irregular with a token frequency of 2. This difference was even more profound in the simulations of P&H; four different irregulars were presented per trial with token frequencies of 4, 3, 2 and 1, and 12 regulars were presented one time per trial. In addition, 10 different sets of 12 regular pairs were presented throughout the course of the simulation. Therefore, over ten trials in each of the P&H simulations, the type frequency for the irregulars was 4, while the type frequency for the regulars was 120. In the simulations of F&E, one irregular and one regular pair were presented three times per trial, and two irregulars and 9 regulars were presented once per trial. Not only is the proportion of low to high frequency pairs much lower than in the previous simulations (10 LF tokens:6 HF tokens for F&E compared to 17 LF tokens:2 HF tokens for the "rule of 78", and 120 LF tokens:10 "relatively" HF tokens for P&H), but the ratio of regulars to irregulars is also much lower (10 regular types:3 irregular types for F&E compared to 17 regular types:1 regular type for the "rule of 78" and 120 regular types:4 irregular types for P&H). It seems therefore, that the preliminary F&E simulations, which coded the whole response, did not show much OR because not enough regular pairs were presented to make them truly "regular" and, more importantly, frequency effects were too strong to be overcome. Coding only the last digit does not increase the number of regulars presented, but it does furnish the model with something similar to the previous experience which allowed the F&E subjects to almost immediately recognize that they only needed to attend to the last digit in their responses, and it reduces the strong frequency effects by eliminating the influence of the first two digits in the response. In essence, by not coding the first two digits of the responses in the F&E simulations, the individual differences between the pairs were discarded, thereby resulting in amplification of the effects provided by the last digit.

Glushko (1979) and Jared et al., (1990) showed that naming, traditionally considered rule based, is affected by neighborhoods. The F&E experiment was conducted in order to determine if neighborhood effects are seen in another system which can be considered rule based. One of Jared et al., major findings was that words whose neighborhoods were composed of high frequency friends and low frequency enemies were named more rapidly than words with low frequency friends and high frequency enemies (refer to Figure 30). The naming task used by Jared et al., assumed that their subjects already knew the words they had to name (that is, familiar words were used). The subjects in the F&E experiment, on the other hand, had to first learn this particular stimulus-response set. Therefore, only the last 10 trials for each F&E subject, when the subjects have presumably had a chance to acquire the set, can be compared in a meaningful way to Jared et al., results. Figure 32 shows that the F&E subjects, like Jared et al., subjects, identified items with high frequency friends and low frequency enemies (HFF-LFE) with greater accuracy than pairs with low frequency friends and high frequency enemies (LFF-HFE). In addition, the F&E simulations also showed this effect (see Figure 38). The difference between the HFF-LFE and LFF-HFE pairs in the F&E experiment was not as large as the difference seen in Jared et al.,. However, this is most probably due to the fact that the F&E subjects did not have time to completely master the stimulus-response set and were therefore more affected by the learning process than Jared et al., subjects. This was not a factor in the F&E simulations since the simulations were run until perfect performance was achieved.

Since the stimulus-response set used in this experiment is based on the "rule of 78", and since the "rule of 78" has been used as an analogy to past tense acquisition, a number of other analyses, related to past tense acquisition, were performed. Prasada, Pinker and Snyder (1990, as cited in Daugherty and Seidenberg, 1992) presented evidence that the frequency of past tense words affects the production of irregular, but not regular past tense forms. According to the traditional theory of past tense acquisition, regular past tense forms are generated by the application of the rule and therefore should not be subject to frequency effects. Irregular past tense forms are formed by looking up the irregular in a memorized list (according to the traditional theory) or generated by a connectionist network (according to a modified traditional theory put forth by Pinker, 1991). Either method of generating the irregular past tense form should be affected by frequency. A one process theory of past tense acquisition, such as SDM, on the other hand, would predict that frequency affects the production of both regular and irregular past tense forms. Figure 33 indicates that over all trials in the F&E experiment, the error rates for the irregular pairs were affected by frequency, while the regular forms were not. However, if only the last ten trials for each F&E subject are examined, frequency does affect the error rates of the regular pairs[38] (see Figure 43). It would therefore appear that, in contrast to Prasada, Pinker and Snyder's (1990) naming latency data, there is at least preliminary evidence that subjects show evidence of frequency effects for regular items, in terms of error rates. As expected, the F&E simulations showed that frequency affects both regular and irregular pairs (see Figure 39).

The work of Jared et al., (1990) and Glushko (1979) has suggested that in naming tasks, words can be classified three different ways, based on how they are generated. Regular consistent (CON) words are named by applying a rule and their neighborhood consists only of other words which follow the rule. Regular inconsistent (INC) words are those which are pronounced using the rule, but they have irregular exceptions in their neighborhoods. Finally there are irregular words which are exceptions to the pronunciation rules. A similar framework for classifying verbs, based on whether or not they follow past tense rules and neighborhood composition has been proposed by Seidenberg (in press). A traditional account of past tense acquisition can only account for two classes of verbs, regular and irregular.

---

[38] For the last 10 trials, a two way ANOVA with repeated measures revealed a significant frequency effect, $F(16,1) = 10.58$, $p<.005$. No regularity effects or interaction were evident. Planned comparisons indicated that high frequency (HF) irregulars had fewer errors than low frequency (LF) irregulars, $t(16)= 3.132$, $p<.01$ and HF regulars showed significantly fewer errors than LF regulars, $t(16)= 2.201$, $p<.05$.

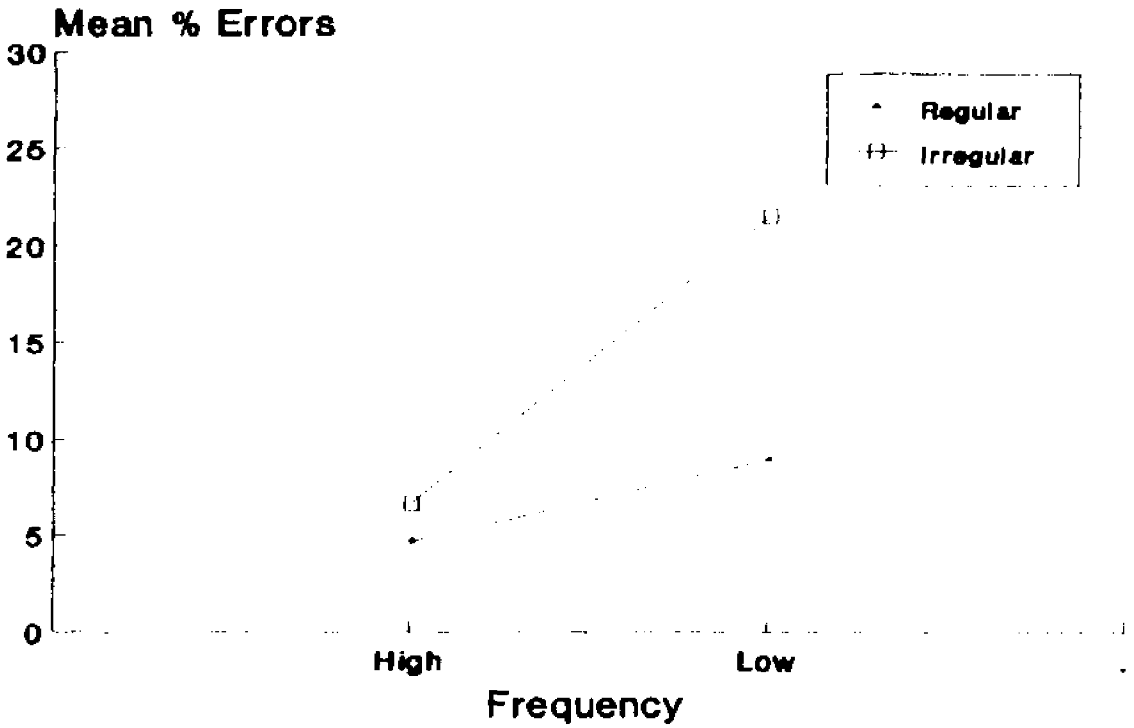


Figure 43. Friends and Enemies (F&E) experiment: Frequency effect. Mean percent errors during the last 10 trials for regular and irregular pairs, based on the frequency of presentation of each pair.

One process theories, like SDM and the connectionist networks, predict all three. In addition, one process systems would predict that regular INC past tense forms should show more errors than regular CON words because their performance is negatively affected by the irregular verbs in their neighborhood. In the F&E experiment, there was no evidence of a distinction between regular INC and regular CON pairs over all trials, since their error rates were not significantly different (see Figure 34). This result would seem to support the traditional account. However, the F&E simulations indicated that fewer errors were made on the regular CON pairs than on the regular INC pairs, supporting the one process theory (refer to Figure 40). This discrepancy between the F&E experiment and the simulations is the most critical inconsistency encountered thus far. Further analysis of the F&E experiment reveals that if the last 10 trials for each subject are examined (once again in an attempt to analyze the trials where at least some mastery of the task is evident and permitting a more reasonable comparison to Jared et al., results), a difference between the error rates of the regular INC and regular CON pairs begins to become evident[39] (see Figure 44). Whether this tendency towards a difference between the error rates for the regular INC and regular CON pairs would have been borne out in an experiment requiring more trials (and therefore more complete mastery of the task), remains speculative at this point. Thus, in this instance, the simulations did not do a very good job of modelling human behavior.

If U-shaped learning and overregularization in the simulations and the F&E subjects are compared, encouraging similarities are evident. If three simulation trials are considered approximately equivalent to one subject trial, the resulting acquisition curves are almost identical (see Figures 35 for the F&E subjects and Figure 41 for the simulations). Both show a substantial decrease in performance on the irregulars immediately following irregular criterion. Both show an increase in performance on the irregulars following the lowest point in the U and both indicate consistently increasing performance for the regulars (with no evidence of U shaped learning). The only difference between the two acquisition curves is in the

---

[39] During the last 10 trials of the F&E experiment, there was an overall effect of consistency, $F(2,16) = 5.38$, $p<.05$, just as in the analysis of all trials. In addition, for the last 10 trials, the mean error rate was 9.71% for the regular INC pairs and 7.84% for the regular CON pairs. This difference approaches significance ($t(16)= 1.572$, $p<.20$) to a greater extent than did the same planned comparison for all trials ($t(16)= 0.532$, $p>.20$) where the mean error rate was 16.75% for the regular INC pairs and 15.71% for the regular CON pairs.

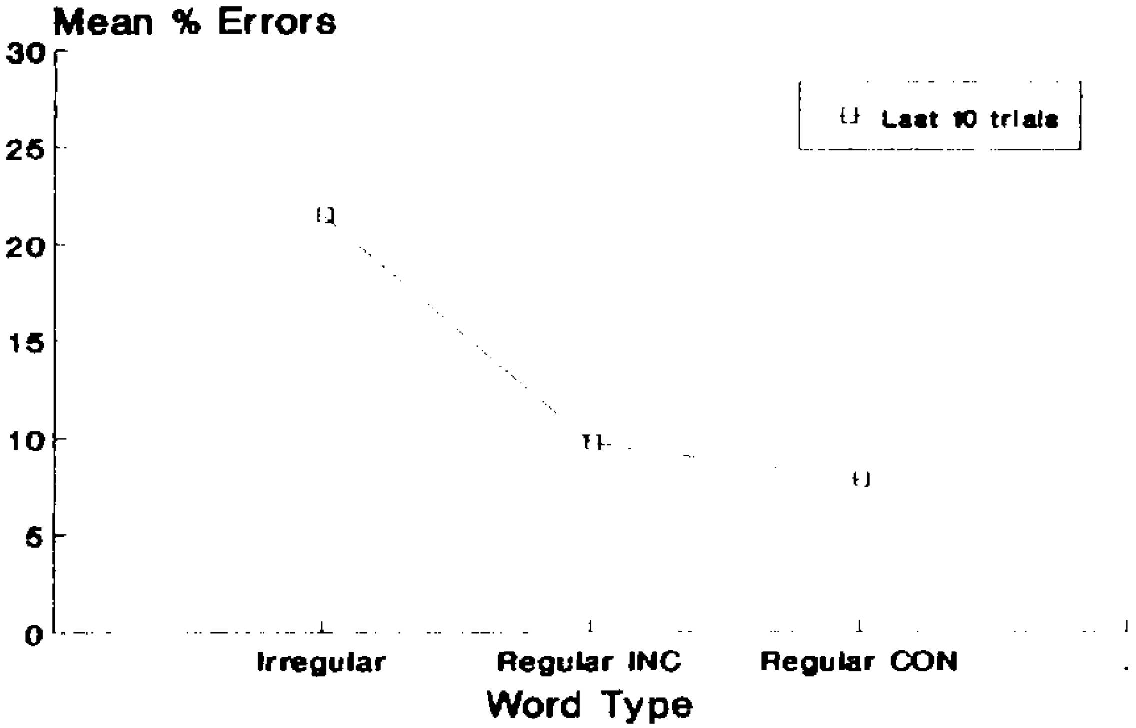


Figure 44. Friends and Enemies (F&E) experiment: Consistency effect. Mean percent errors during the last 10 trials for irregular, regular inconsistent (INC) and regular consistent (CON) pairs.

right hand arm of the U. The simulations returned to perfect performance for the irregulars, while the F&E subjects only achieved a level of 90% by the end of the experiment. This difference may be explained rather trivially. In all probability, the experiment did not involve a sufficient number of trials to allow the F&E subjects to once again reach perfect performance on the irregulars.

Finally, it seems prudent to summarize how well the simulations modelled the behavior represented by the F&E experiment:

1) Both the simulations and the F&E subjects made fewer errors on the irregular with high frequency friends and low frequency enemies than on the irregular with low frequency friends and high frequency enemies. This lends support to the notion that in both human memory, and SDM, past tense generation is affected by neighborhood composition.

2) Frequency was shown to affect irregular pairs in both the simulations and F&E. Frequency affected the regulars in the simulations, but only during the last 10 trials for the F&E subjects. Since there has been little behavioral data thus far to indicate any frequency effects for regular items, this result warrants further investigation.

3) The only major discrepancy between the simulations and F&E was found when consistency effects were examined. While over all trials, the F&E subjects showed no difference in

error rates between regular inconsistent and regular consistent items, an examination of the last 10 trials reveals that this difference approached significance. The simulations showed a significant difference between the error rates of regular INC and regular CON pairs, and therefore did not model the behavioral data for all trials. However, the difference found in the last 10 trials of the F&E experiment suggests that the subjects may have shown an effect had they been run longer, and if that was the case, SDM is clearly capable of simulating the effect.

4) Finally, U-shaped learning, accompanied by an increase in OR rates, was evident in both the simulations and F&E.

III. General Discussion

One reason that the acquisition of past tense English verbs has received considerable attention recently is the hope that examination of this domain will provide theories which will lead to a more general understanding of cognitive processing. Traditionally, a modular view of past tense acquisition has been supported. The traditional theory proposes that two distinct systems are required to account for the inflection of regular and irregular verbs. The past tense of irregular verbs are thought to be memorized and therefore stored as individual entries in the mental lexicon, while the inflection of regular verbs is believed to occur on line by implementing a rule. This view has been supported by behavioral evidence which indicates that children acquire the irregular past tense, but not the regular past tense, in a U-shaped developmental sequence. The traditional theory accounts for these data by arguing that at first, children produce only correct past tense forms (whether regular or irregular) because they are just memorizing them. Once the child begins to extract the past tense inflection rule, through experience with regular verbs, he or she begins to overregularize irregulars, resulting in a decrease in correct performance on the irregulars, compared to early performance. As described previously (see section I.B.1) the final stage of development, a return to near perfect performance on the irregulars, is not clearly explained by the traditional theory.

In 1986, Rumelhart and McClelland published an account of past tense acquisition in a PDP network, which was able to simulate a number of aspects of past tense acquisition, including U-shaped learning and overregularization. According to R&M this system did not specifically encode the past tense inflection rules, demonstrating that a single process system could account for the behavioral data, and challenging the traditional two process theory. This model was extensively criticized, but subsequent connectionist models (MacWhinney and Leinbach, 1991; Plunkett and Marchman, 1991 & 1993) have made similar claims using networks and learning algorithms designed to eliminate some of the features which were criticized in the R&M model.

Although none of the above mentioned connectionist models perfectly simulate the behavioral data, and all are subject to considerable criticism, the debate sparked by their results has prompted the modification of the traditional two process theory. One of these modifications, the Marcus et al., (1992) "blocking and memory retrieval failure" theory, was described earlier (refer back to section I.B.2). The second was proposed by Pinker (1991). Pinker suggests a theory which includes a computational component for rule governed regular verbs, and an associative memory system for the irregulars. This associative memory system, unlike a simple memorized list of irregular stem-past pairs, has properties similar to the connectionist networks. He suggests that regulars are still produced on-line by a rule which concatenates the regular suffix to the base of the stem. Irregulars, on the other hand, may be thought of as memorized pairs of words with linkages between the pairs stored in an associative memory structure. These irregular pair associative linkages facilitate the production of irregular past forms which belong to the same "family" (e.g., sling-slung and sting-stung). All in all, although these two theories take considerably different approaches, both stipulate that two processes are required to account for acquisition of regular and irregular verbs.

The debate between the one and two process theories is further complicated by the introduction of SDM (Kanerva, 1988). As an associative memory model, SDM more closely resembles a standard random access memory model than a neural network (Denning, 1992). However, although the architecture is considerably different, the simulations reported as part of this work indicate that SDM, like the connectionist networks, may provide an alternative to the two process theories of language acquisition. SDM does not lend itself to the encoding of specific rules since it stores distributed representations of its input and allows for retrieval through statistical reconstruction of the contents of numerous memory locations. Therefore, since the claim is being made here that SDM is able to mimic human behavior, the view that two processes may not necessarily be needed to account for language acquisition, or cognitive processing in general, is strengthened.

Before addressing how well SDM was able to simulate human behavior, comparisons will be made between SDM and the aforementioned neural network models of past tense acquisition. In terms of network architecture and training schedules, SDM has a number of advantages over the neural network models. First, all four network models (R&M, L&B, P&M-91 and P&M-93) rely on error calculation, while SDM does not. Since error calculation can be viewed as a type of negative feedback, this distinction is particularly meaningful since there is little evidence indicating that children receive any systematic negative feedback when they are learning the past tense (Kuczaj, 1977; Marcus et al., 1992). Secondly, and more importantly in terms of the theoretical implications of each model's results, the SDM simulations used a continuous input training set, while both the R&M model and the P&M-93 model used discontinuous training regimes.

Thirdly, McCloskey and Cohen (1989) and Ratcliff (1990) both claim that distributed network models using the back-propagation algorithm are subject to "catastrophic interference". While network simulations are typically trained concurrently (that is, all items in the set to be learned are repeatedly presented during a particular simulation), McCloskey and Cohen point out that humans often learn sequentially (for example, addition is typically learned, and pretty well mastered, before multiplication learning takes place). McCloskey and Cohen and Ratcliff tested different distributed neural networks to see how they handled sequential learning problems. Both concluded that new learning may interfere catastrophically with old learning, when the networks were trained sequentially. Retroactive interference (RI) in and of itself is not the problem, since humans often exhibit RI. However, the magnitude of interference observed in the network simulations far exceeds that observed by humans. As

McCloskey and Cohen flippantly comment "the magnitude of the observed interference (in the network) makes it seem more like retrograde amnesia than retroactive interference". McCloskey and Cohen and Ratcliff both vary numerous parameters, including the number of hidden units, the strategies used to modify the weights of the hidden units, the learning rate parameter, the amount of training given to the network on the original training set, and the target activation levels of the output units. Although some of these variations produced less interference, none of the variations produced interference levels anywhere near as low as that exhibited by human subjects[40]. Both Ratcliff and McCloskey and Cohen place the blame for this interference on the perceptron learning algorithm (for the R&M model) and back-propagation (for all other models). Since these learning procedures require that each presentation of an item to be learned results in the adjustment of many unit weights in the network, a set of items is only learned when a very distinctive pattern of weighted units is achieved.

---

[40] For human subjects, unlearning of the first set of items in the classic retroactive interference experiment rarely exceeds 50% (Barnes & Underwood, 1958 as cited in McCloskey & Cohen, 1989). However, regardless of the parameters used in the network simulations of this classic RI experiment, McCloskey & Cohen report that, following training on the second set, the best performance achieved on the first list was 31% (or 69% interference). Even more startling, in the vast majority of simulations, first set performance was reduced to virtually 0% correct by second set training sufficient to yield only 25% correct on the second set.

The success of the model, in identifying the items in the training set, is dependent on this pattern of weighted units. When a new set of items is presented to the same network, the weights of the units are modified to accommodate the new items and the original unique pattern is drastically disrupted, resulting in interference. According to McCloskey and Cohen, it is unlikely that the back-propagation algorithm can be modified to protect previously learned patterns, since it does not store specific patterns, but instead induces from training a function that will map an input pattern to an output pattern. Therefore, since it cannot identify particular patterns, it cannot protect previously learned patterns. In other words, because the network has no representation of any of the learned patterns "as a whole", it cannot identify and protect previously learned patterns from interference. SDM, although it does store items in a distributed manner, does store the pattern "as a whole" and is therefore less susceptible to interference. Theoretically, when SDM is trained sequentially, only the memory locations within the specified storage radius of the new items will be affected. In other words, only the previously learned items which are located in memory locations within the storage radius of the new items, should be affected by interference. This should result in much more local, and therefore less than catastrophic, interference. Therefore, while distributed network models using back-propagation are subject to catastrophic interference during sequential learning, SDM undergoes a more graceful degradation.

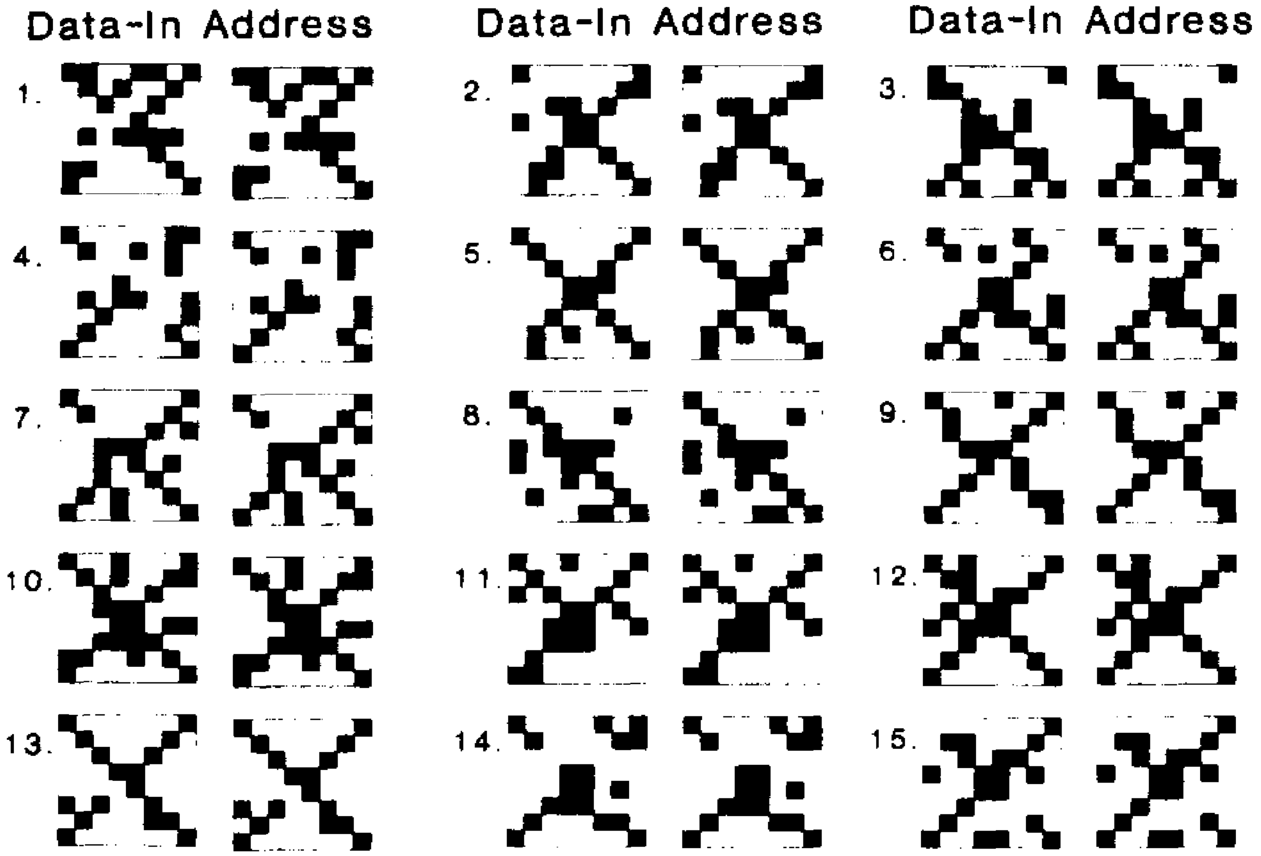
Evidence of the graceful degradation properties of SDM can be seen in the simulations reported here. Although all the responses, in each of the three sets of simulations, were initially stored at "noisy" stimulus addresses, SDM was eventually able to learn all the stimulus-response patterns it was presented with. Even in the beginning of each simulation, when noise levels were very high, SDM was able to retrieve the correct response some of the time. A little extrapolation from these results provides insight into how SDM should handle retroactive interference. Consider the possibility of one of the SDM "rule of 78" simulations as the initial training set in a RI simulation. If the memory resulting from this initial training is then presented with a new training set, it is likely that the two training sets would address at least some of the same memory locations, resulting in their degradation. However, unlike the connectionist models, this input should not "catastrophically" interfere with SDM's ability to extract the originally stored items. Based on the evidence that initially, when presented with very noisy (which can be considered extremely degraded) stimuli, SDM could retrieve the correct "rule of 78" response at least some of the time, SDM would probably exhibit a decrease in the level of performance for the originally stored items, but the presentation of the new training set should not prevent SDM from being able to retrieve the originally stored items. The amount of interference seen should be proportional to the number of memory locations which are addressed by both training sets, and the extent to which those locations are degraded. In fact, if the second training set was sufficiently different from the "rule of 78" training set (for example, if the P&H stimulus-response set was used as the second training set), it is entirely possible that no interference would be seen, since inputs which are very different (that is, inputs which are separated by a large Hamming distance) could be stored at memory locations which are distant from each other in the memory space (as they are in the "rule of 78" and P&H simulations). Therefore, in situations involving RI, SDM's graceful degradation properties should allow it to perform more like humans than the connectionist models.

Thus far, SDM's graceful degradation properties provide the first clue that SDM is capable of simulating more general human behavior, rather than just being able to simulate the much more specific situations addressed by the three sets of simulations reported in the main body of this work. In addition, Denning (1989) provided evidence that, on an elementary level, SDM is able to simulate both human knowledge of temporal relationships (Denning, 1989, Sequencing figure, p. 335) and the human ability to abstract a prototypical category representation from a set of exemplars (Denning, 1989, Abstraction figure, p.334). A simple demonstration, using the implementation of SDM employed in this work, further illustrates SDM's ability to abstract a prototype. Figure 45 presents the abstraction of the letter X using SDM as an autoassociative memory system. Fifteen exemplars of the letter X were generated by randomly reversing 15% of the bits in a prototypical X (Figure 45A - Data_In 1-15). These exemplars were stored at addresses which matched their input data. That is, the data vector and the address vector used to store the data were the same (Figure 45A - Address 1-15)[41]. Following storage of all 15 patterns, data were retrieved from an address which represented yet another "noisy" X (Figure 45B - Address 1), resulting in a pattern which differed from the prototype by only 1 bit (Figure 45B- Data Out 1). The data were then read from the address of this "nearly perfect" X (Figure 45B - Address 2), and the "prototypical X" was retrieved (Figure 45B - Data_Out 2). This "prototypical X" was stable, since subsequent retrieval from this address produced the same results. It is therefore clear that, like humans, SDM is not only able to extract a signal from noise, but it is also capable of forming a prototype from a set of examples.

---

[41] This simulation used a randomly generated seed, 64 bit vectors, 2000 memory locations and storage and retrieval radii of 20.

A. Training:



B. Testing:

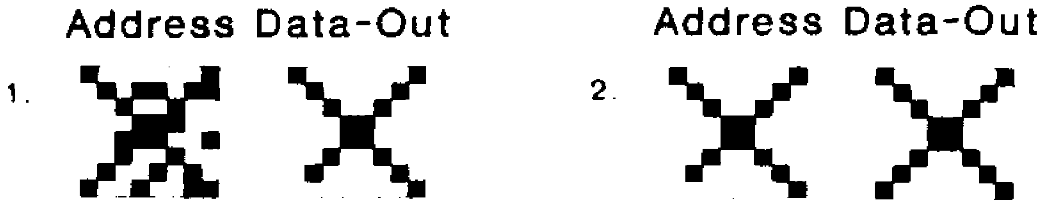


Figure 45. SDM simulation: Abstraction of a prototypical "X" from a set of 15 noisy exemplars. In this example, SDM was used as an autoassociative memory system, that is, the data were stored at addresses which matched the input data.

While simple demonstrations of SDM's ability to simulate some very basic human cognitive abilities are important in the evaluation of SDM as a model of general human memory, the principal goal of this work was to test SDM's ability to mimic more specific human cognitive abilities. SDM's success in the "rule of 78" simulations, the P&H simulations, and the F&E simulations will be evaluated by comparison to: a) the behavioral data, and b) the results of simulations performed by other researchers who have used neural network models (R&M, M&L, P&M-91 and P&M-93, among others).

Before attempting any evaluation of the SDM simulations, a discussion of one SDM parameter must be addressed. The results of all the SDM simulations in this work were dependent on using a strategy of decreasing noise over the course of the simulations. Initially, pilot simulations were run with constant noise values throughout each simulation, and the results did not show U-shaped learning or overregularization. Therefore the decreasing noise strategy must be considered critical to the reported simulation results. In this respect, the implementation of SDM used in this work is actually a modification of Kanerva's (1988) SDM model. While this strategy was uncovered through trial and error, once discovered, its behavioral justification became apparent. Human acquisition, more often than not, occurs under less than ideal conditions. When acquiring the past tense of English, for example, a child is burdened with a number of intrinsic and extrinsic distractors. As alluded to previously (see section II.A.3), the decreasing noise parameter used in these SDM simulations may be considered analogous to the noise experienced by children during the acquisition process. In children, extrinsic noise probably plays the greatest role and may take the form of competition from other environmental factors. A young child's world is full of things that can distract him or her from the task at hand (language acquisition in this case). Children do not attempt to learn language by devoting their attention to the task exclusively. Instead, they assimilate language during the course of their everyday activities and therefore their attention may be diverted by many things, including: 1) non-verbal interaction with their parents and other individuals they encounter, 2) play activities which stress motor abilities over language skills and, 3) physical considerations like hunger and comfort. These types of distractors are particularly relevant in language acquisition, since children have relatively short attention spans during the developmental period in question. In addition, intrinsic factors may also play a role. The human cognitive system, since its basic components are neurons, always experiences some degree of intrinsic noise in the form of spurious neural activity. This is particularly true for children, since neural fibers undergo myelination throughout childhood. It therefore seems reasonable to assume that language acquisition occurs under noisy conditions. The decision to use a strategy of decreasing noise in the simulations was reached by assuming that as children become more proficient with language, extrinsic noise (and perhaps intrinsic noise) has a less critical effect on the learning process. This assumption seems reasonable when one considers that as children get older, their attention spans increase.

Rumelhart and McClelland (1986) considered the "rule of 78" a simple analogy to past tense acquisition, since it consists of a number of pairs which follow a rule, and includes an exception to this rule. Before implementing simulations of past tense acquisition on a more comprehensive scale, R&M explored their PDP model's ability to handle this simple system. A similar strategy was used in this work. The "rule of 78" was simulated using SDM before attempting to simulate the more complex cognitive abilities represented by the Palermo and Howe (1970) study and the friends and enemies experiment. Since the "rule of 78" is somewhat analogous to past tense acquisition, the results of the SDM "rule of 78" simulations will not only be compared to R&M's "rule of 78" simulations, but they will also be compared to: 1) Rumelhart and McClelland's (1986), MacWhinney and Leinbach's (1988), Plunkett and Marchman's

(1991) and Plunkett and Marchman's (1993) simulations of past tense acquisition and 2) the child acquisition data.

One methodological consideration should be addressed before evaluating SDM's ability to learn the "rule of 78". A decision was made to present the irregular twice per trial in these simulations. This decision was made for two reasons. First, in English, irregular verbs tend to be higher frequency than regular verbs. It therefore seems reasonable to present the irregular more often than the regulars. However, the decision to present the irregular twice per trial, as opposed to three or four times, was an empirical one. If the irregular was presented more frequently, irregular performance in the simulations was almost immediately perfect, with little evidence of U-shaped learning or overregularization. Although this may be seen by some as a methodological "trick" to ensure success, there is behavioral support for the decision. Palermo and Howe's (1970) experimental analogy to the past tense showed that frequency effects play a major role in human acquisition. The irregular which was presented four times per trial showed very little OR, and a very shallow U, while the irregular presented once per trial[42] showed much more OR and more profound U-shaped development. In addition, the

---

[42] Although this irregular was only presented once per trial in the P&H experiment, it was still more frequent than the regulars, since on average, each regular was only presented once every ten trials.

eleven irregular verbs used most frequently by Adam, Eve, Sarah and Abe (Marcus et al., 1992 Tables A5-A8, p.148-151), excluding "be" and "have" (which can also serve as auxiliary verbs, thereby inflating frequency measures) are, on average, 2.4 times as frequent as the eleven most frequent regular verbs listed in Francis and Kucera's, 1982 rank list[43]. Therefore, rather than being viewed as a methodological device which only serves to guarantee the desired results, this decision should be viewed as a method of making the simulations more closely match the behavioral evidence.

Returning to the evaluation of the simulations, a strong case may be made for the view that the SDM simulations of the "rule of 78" embodied all major aspects of U-shaped learning. Subsequent to achieving 100% correct performance early in the simulations, the irregular showed decreased performance, accompanied by overregularization. During this period, the irregular past tense form coexisted with the overregularized incorrect form. The final aspect of U-shaped development, a return to perfect performance for

---

[43] The frequency measures used in this comparison are the frequencies for ALL forms of these 22 verbs. Since Marcus et al., do not list the regular verbs used by Adam, Eve, Sarah and Abe, it was necessary to take the eleven most frequent regular verbs from Francis & Kucera's rank order list. However, since 7 of the 11 irregular most frequent verbs used by the children were ranked higher than the most frequent regular verb by Francis & Kucera, this list of high frequency regulars should at least approximate that used by the children.

the irregular, was also observed. In contrast, the regulars showed a gradual and steady increase in performance over the course of the simulations, with no evidence of U shaped learning. This developmental sequence was achieved without the use of a discontinuous input set, which gives it a clear advantage over R&M's "rule of 78" and past tense acquisition simulations. Therefore, since there is little behavioral evidence for drastic discontinuity in the child acquisition literature, the SDM simulations more closely resemble the child acquisition data than do the R&M simulations. In addition, although M&L and P&M-91 also used a continuous input set, they were unable to simulate early correct usage in their simulations. Of course, the point may be made that the SDM simulations did not show 100% correct performance at the very beginning of the simulations, but as P&M-91 point out, children start out with at least some knowledge of the properties of language before they begin to learn the past tense, while computer simulations start with no a priori information. Therefore the initial few trials of the SDM simulations may be considered analogous to children's early experience when they gather preliminary information about language.

In these simulations, the point at which each simulation first reached 100% correct on the irregular was considered the beginning of the U-shaped developmental sequence (early correct usage), and for all 10 simulations, this perfect performance was not achieved for an extended period of time (usually only 1 or 2 trials). This result is bound to provoke criticism. However, if the behavioral concept of "early correct usage" is examined more closely, this potential criticism loses some of its potency. In effect, the behavioral evidence for early correct usage of irregular past forms is misleading. What the child data actually show is an early period of "non-overregularized" irregular past tense usage. Marcus et al., (1992) have conducted one of the most in depth analyses of past tense acquisition, drawing data from numerous sources. When they graph the U-shaped development of Adam (Kuczaj, 1977) and the children in the Brown (1973) study (Eve, Sarah and Adam) they label the x axis as "% irregular past tenses correct", but they are actually plotting 1-OR rate (e.g., Figure 12a). An examination of the OR curves of most of the individual SDM simulations also reveals a more extended period of "early correct usage" using this type of graphical representation (see Appendix A, Figures AP-A1 through AP-A10). Marcus et al., only examined past tense usage in "obligatory" past tense contexts. However, early in development, because of the child's limited experience with language, obligatory past tense contexts may not be readily apparent. In addition, during this early stage of past tense acquisition, children may not mark verbs at all, even if use of the past tense is indicated semantically. Marcus et al., excluded these "unmarked" errors from their analysis of U-shaped development, further reducing the number of possible early errors. Therefore, had all behavioral errors been taken into account, an actual extended period of early irregular past tense usage seems unlikely.

Another factor which must be considered becomes apparent if irregular performance before and after the first trial at 100% correct on the irregular is examined in the SDM "rule of 78" simulations. In the trials before this criterion point is reached, the irregular performs, on average for all 10 simulations, at a rate of 47.6% correct. In the 10 trials following the first trial at 100% correct, on average for all 10 simulations, the irregulars perform at 81.6% correct[44]. This difference is highly significant, $t(9) = 9.447$, $p<.001$. Therefore, it is clear that although the simulations do not show an extended period of perfect performance on the irregular in the trials immediately following the first irregular trial at 100% correct, they do show evidence of a much higher level of performance than before this criterion is reached. In other words, following this point, there is clear evidence that the simulations have mastered the irregular pair.

Just as there are those who may question the U-shaped learning curve obtained in these SDM simulations, Virginia

---

[44] It should also be noted that none of the simulations showed better irregular performance before criterion than they did after criterion.

Valian (personal communication, December 14, 1993) has suggested that U-shaped learning in general (that is, in language acquisition as well as in any simulation of U-shaped learning) may only be the result of sampling error. In the language acquisition literature, this argument has been addressed by Marcus et al., (1992). They have tested the hypothesis that there is actually a fixed rate of overregularization throughout the developmental period in question and that ORs are just spurious sampling errors. Since the children in the language acquisition literature have often not been observed for an extended period of time and there is therefore little data available for the existence of the right hand arm of the U (that is, a return to perfect performance), Marcus et al., have confined this analysis to the left hand arm of the U (early correct performance followed by a period of OR). If the OR rate is p, then under the null hypothesis that the OR rate is fixed throughout the period under examination, the chance of the first irregular verb form being correct is $1 - p$, the chance of the first two irregular occurrences being correct is $(1 - p)^2$, and so on. Therefore, Marcus et al., tested whether there was an improbable run of correct irregular past tense forms at the beginning of the children's records by calculating $(1 - p)^n$, where n is the number of irregulars in the child's transcript preceding the first OR error. For three of the seven children with enough data to do such an analysis, the null hypothesis of a fixed OR rate can be rejected since the probability of such an occurrence was less than 0.001. However, for the other four children considered individually, no conclusion could be made since the probabilities ranged between 0.255 and 0.659 (see Marcus et al., 1992, Table 3, p. 42). Marcus et al., then treated the seven children as a single sample using a meta-analysis and determined that the overall probability of obtaining the data under the null hypothesis was less than 0.001, suggesting that U-shaped learning cannot be attributed to sampling error.

Table 12 shows the results of this analysis applied to the SDM simulations of the rule of 78. Since the SDM simulations of the rule of 78 showed evidence of the right hand arm of the U-shaped curve (that is, a return to perfect performance) while the children analyzed by Marcus et al., did not, the OR rates used in this analysis were calculated only over those trials before irregular performance stabilized at 100% correct. Unlike the analysis of the child data, the null hypothesis of a fixed OR rate cannot be rejected for any of the nine SDM simulations of the rule of 78 considered individually. In addition, if these 9 simulations are treated as a single sample using the same meta-analysis as that used by Marcus et al.,(1992), the null hypothesis still cannot be rejected, since the overall probability of obtaining the data under the hypothesis of a

Table 12

Tests of U-Shaped Development: SDM Simulations of the Rule of 78[45]

Simulation	First OR Trial	Consecutive Correct in Preceding Trials	OR Rate	Probability
DX	7	10	.0341	.7068
PX	12	29	.0208	.5436
RX	3	4	.0424	.8409
TX	7	10	.0300	.7374
UX	4	3	.0244	.9286
VX	4	7	.0435	.7325
WX	7	19	.0244	.6254
YX	4	5	.0385	.8218
ZX	3	2	.0489	.9046

---

[45] Only 8 of the ten SDM simulations of the rule of 78 are included in this table since two of the simulations had an overregularization rate of 0.

fixed OR rate is 0.4784. Therefore, the possibility exists that in the SDM simulations of the rule of 78, the observed U-shaped curve may be due to sampling error. However, it should be noted that the analysis done by Marcus et al., (1992) only included seven children, and the null hypothesis of a fixed error rate could only be rejected for three of them. In addition, the OR rates for these seven children included data from many different irregular verbs, while the SDM simulations of the rule of 78 had only one irregular pair. Had these simulations used a larger number of irregulars, perhaps the results would have been more similar to the behavioral data. Nevertheless, future work with SDM must address this issue.

Even though this stringent test of U-shaped learning is inconclusive for the SDM simulations of the "rule of 78", these simulations do show low levels of overregularization. Although early reports of the phenomenon suggested that overregularization was prevalent, more recent investigations (most notably Marcus et al., 1992) have shown OR rates to be very low (median OR rate = 2.5%, mean OR rate = 4.2%). Although the 10 SDM simulations reported here had median OR rate (1.4%) about half that of the behavioral data, and a mean OR rate (1.2%) 70% less than that reported in the child data, these quantitative differences can easily be accounted for. First, the language acquisition rates represent many irregular verbs, while these simulations only included one irregular pair. Marcus et al., provide evidence that different verbs have different OR rates and in fact, the SDM simulation results look very similar to the U-shape learning of one particular verb examined in isolation (see Figure 11a&b). Secondly, the pilot simulation (ZX) which was run with more noise indicates that OR rates more similar to the behavioral data can be attained by varying only one SDM parameter (see Table 2).

The OR rates for some of the other network models of past tense acquisition do not compare as well to the behavioral data. Although specific OR rates are not given for R&M's "rule of 78" simulation, R&M state that "even after 40 trials, the model still gets the wrong answer on Units 7 and 8 for the (147) pattern more than half the time" (McClelland & Rumelhart, 1986, p.232). This suggests a much higher OR rate than is seen in either SDM or the child data. R&M's more elaborate past tense acquisition simulation also appears to show more overregularization than SDM and the behavioral evidence. During the 200 training trials, the OR rate reaches almost 50% at its peak, and is actually higher than the rate for correct irregular past tense forms at this point (see McClelland & Rumelhart, 1986, Figure 5, p.243, trials 11-30). There is little behavioral evidence that most children produce OR's at a higher rate than they produce correct responses, at any time during the acquisition process[46]. In addition, a very gross estimate over the 200 training trials (made by visual inspection of the above mentioned figure 5) puts the overall OR rate for this simulation at about 15%, which is much higher than children's mean OR rate as reported by Marcus and his colleagues.

Plunkett and Marchman (1991) also do not present an overall OR rate for their phone 34 simulation of past tense acquisition, however, based on the information provided (Table 5, p.91 and Table 6, p. 93), it is possible to roughly estimate[47] that the mean overall OR rate was 5%, which closely matches the mean overall OR rate for children reported by Marcus et al. Similarly, in their subsequent simulation, based on a comparison of Adam's U-shape learning curve to P&M-93's simulation, (see P&M-93, Figure 3), the

---

[46] For the Adam, Eve, Sarah and Abe, the four children examined in greatest detail by Marcus et al (since their data allowed the most in depth analysis), only Eve and Abe showed evidence of an OR rate higher than the correct past tense rate at any time during development (see Marcus et al, 1992, Figure 8, p. 48 for Eve and Figure 10, p. 49 for Abe). However, based on overall OR rates, both Eve and Abe may be considered to exhibit unusually high overall OR rates (13% for Eve and 24% for Abe) compared to the other 23 individual children whose mean overall OR rate was only 3% (from Marcus et al, 1992, Table 2, p. 36).

[47] P&M-91 Tables 5 and 6 present the distribution of error types for epochs 1-15, 20,25,30,40 and 50. Based on their description of error types, an SUF error is considered an OR error for the purpose of estimating the overall OR rate for the simulation. For epochs without listed SUF rates, the epoch was assigned a SUF rate equal to the closest listed epoch's SUF rate (e.g. epoch 46 is not listed and is therefore assigned the SUF rate for epoch 50).

simulation also shows OR rates similar to the behavioral data. However, it should be kept in mind that for the P&M-93 simulation, a discontinuous input set was used during training. This discontinuity is much less drastic than that used by R&M, and their gradual, incremental increase in vocabulary size is supported by parental report measures (Marchman and Bates, in press). However, the validity of the parental report measures used by Marchman and Bates has been questioned by Marcus et al., (1992). According to Marcus et al., Marchman and Bates only presented the parents in their study with a small subset of the English verbs that children of the age groups under examination are known to use. Therefore, according to Marcus et al., Marchman and Bates' vocabulary estimates are underestimated. Thus, it is still not clear that P&M-93's input strategy can be considered comparable to the way children acquire new verbs.

Overall the "rule of 78" simulations have provided strong evidence that SDM is capable of simulating the human phenomena of U-shaped learning and overregularization, and in many ways SDM was superior to the connectionist models implemented by other researchers. In addition, while both R&M's past tense simulations and M&L's simulations have been criticized for not actually being one process systems (because of the Wickelfeature representation used by R&M and the direct input-output link in M&L's network architecture), the same type of criticisms cannot be made of SDM. In these

SDM simulations, neither the coding system, nor the model's architecture, treat regulars and irregulars differently, or incorporate the past tense rule. Therefore, the SDM simulations of the "rule of 78" provide clear evidence that a one process language acquisition system cannot be discounted.

While acquisition of the "rule of 78" may be considered analogous to human learning, it does not represent a situation actually encountered behaviorally. Therefore, SDM's success in this situation is not sufficient to support the view that SDM is capable of mimicking actual behavior. The SDM simulations of Palermo and Howe's (1970) experiment were an attempt to evaluate SDM under conditions which more closely match actual human acquisition. Although comparison to other network models is not possible, since there have been no other simulations of P&H's experiment, SDM's results may be compared to the human data. In addition, since P&H designed their study as an experimental analogy to past tense acquisition, comparisons to the child acquisition data are also appropriate.

The resemblance of the SDM simulation results to P&H's results was rather striking in a number of ways. In both cases, the irregulars were learned well before the regulars. In addition, frequency effects were also evident in both acquisition situations. As the frequency of the irregular increased, the acquisition rate also increased. Both these results correspond well to past tense acquisition in children, since children tend to inflect highly frequent verbs (which are often irregular) earlier than low frequency verbs.

Secondly and more importantly for the purpose of this discussion, both P&H's subjects and the SDM simulations showed evidence of U-shaped learning and overregularization, once criterion was reached on the irregulars. In both cases, the U was rather shallow and the levels of OR were low. Once again, this closely mimics the past tense acquisition data for children. As Marcus et al., (1992), among others, have demonstrated, children begin to show evidence of OR and U-shaped learning after they initially use the irregulars correctly. In addition, children show very low levels of OR, and their U-shaped learning can in no way be considered profound.

Thirdly, if the learning curves (graphed as 100-OR rate as in the Marcus et al., analysis of the child acquisition literature) of both P&H's subjects and the SDM simulations of P&H are examined, they look remarkably similar even though neither the P&H subjects as a group, nor the SDM simulations of P&H taken together show any evidence of U-shaped learning (see figure 46 for P&H's subjects and figure 47 for the simulations). Since behavioral reports of U-shaped learning refer only to data from individual children, no conclusions about U-shaped learning can be made from these aggregate data. However, when the individual P&H subject data and the individual SDM simulations of P&H are examined, it appears as if both the subjects and the simulations fall into one of three categories. Some subjects and simulations do not appear to show any evidence of U-shaped learning as can be seen in figures 48 (for P&H subject #13) and 49 (for simulation BB). Other subjects and simulations show evidence of U-shaped learning after an initial period during which either the subject or the simulation reaches a level of no OR errors. Figure 50 reveals that from trial 2 through trial 14, P&H subject #8 shows an extended period of no OR errors, followed by a period of OR (trials 15-23) and then a return to perfect performance (at least in terms of no evidence of OR errors). Similarly, figure 51 indicates that SDM simulation CC showed no evidence of OR errors from trial 20 through 44, followed by a short period of OR (trials 45-46) and then a return to an extended period of perfect performance (trials 47-500). Finally, P&H subject #3 and SDM simulation VX are representative of those subjects and simulations that appear to show what could be classified as classic U-shaped learning, that is, an initial period of no OR errors, followed by a period of OR and then a return to perfect performance (see figure 52 for P&H subject #3 and figure 53 for SDM simulation VX). Therefore, SDM was able to simulate all three learning patterns exhibited by P&H's subjects.

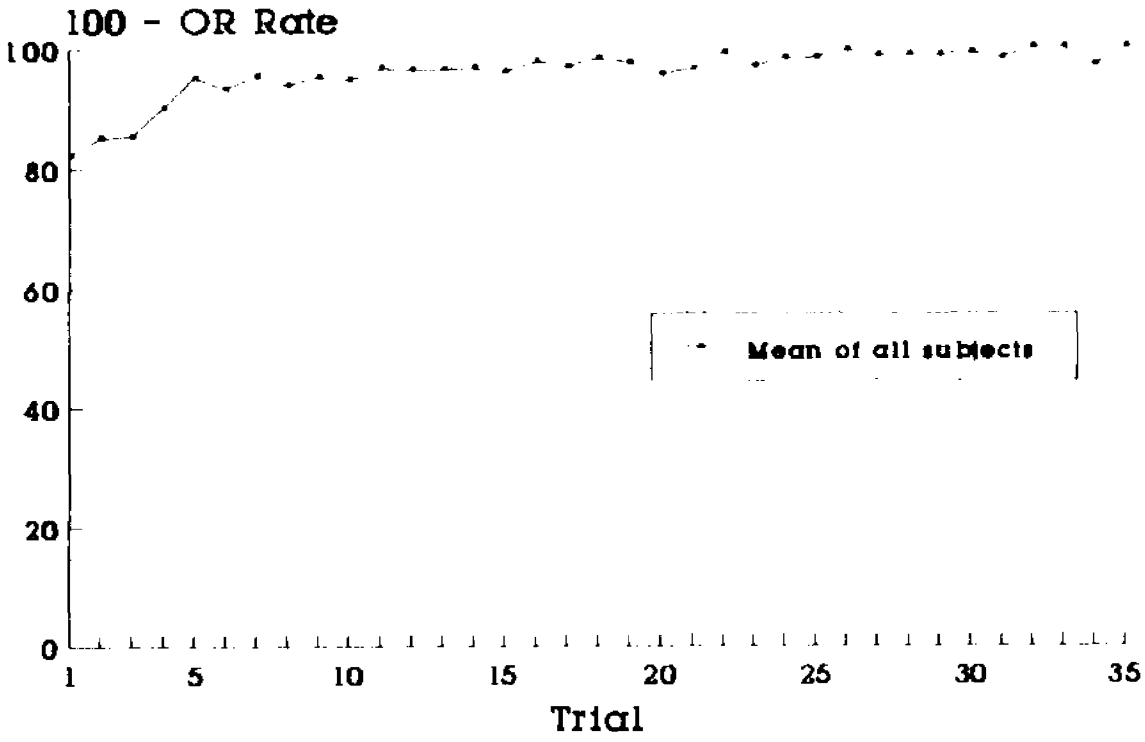


Figure 46. Palermo and Howe (1970): Mean learning curve for all subjects.

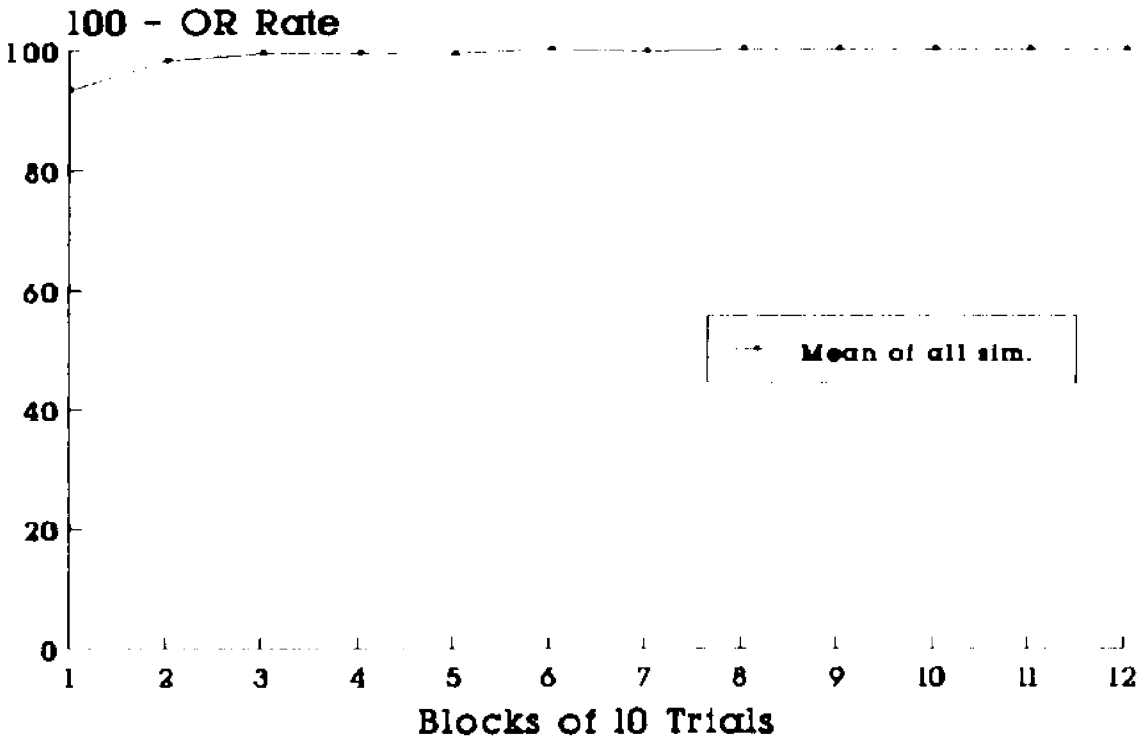


Figure 47. SDM simulations of P&H(70): Mean learning curve for all simulations.

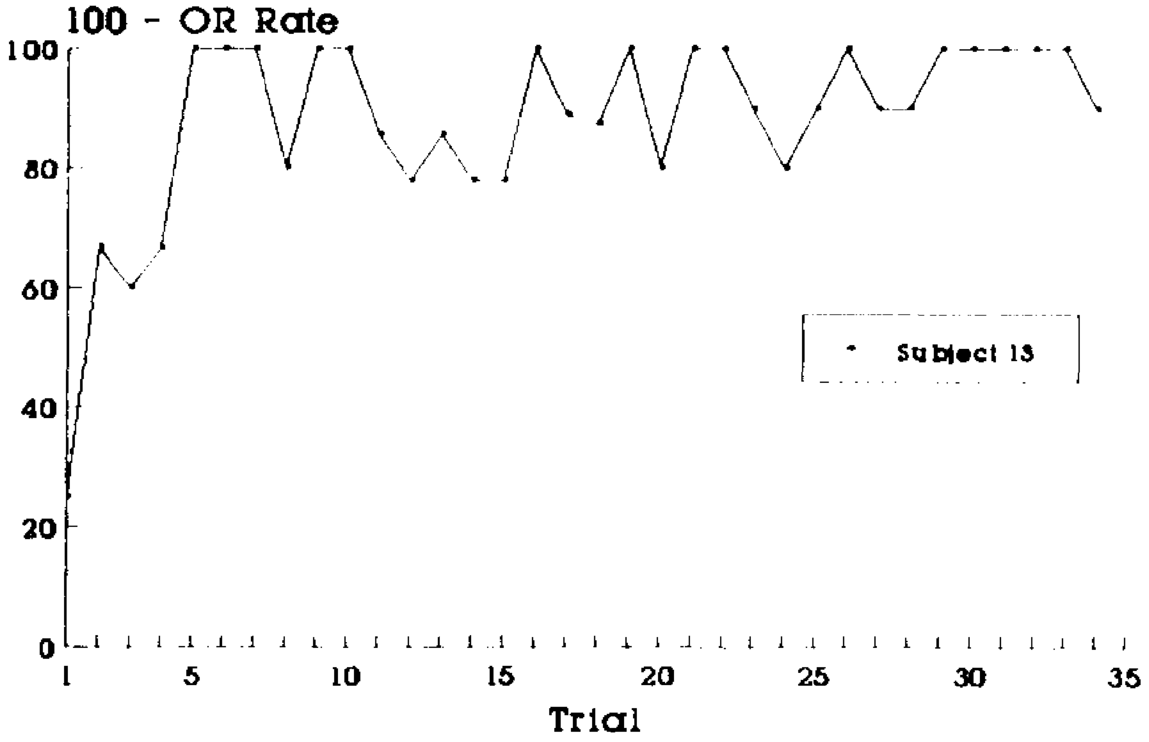


Figure 48. Palermo and Howe (1970), Subject #13: No evidence of U-shaped learning.

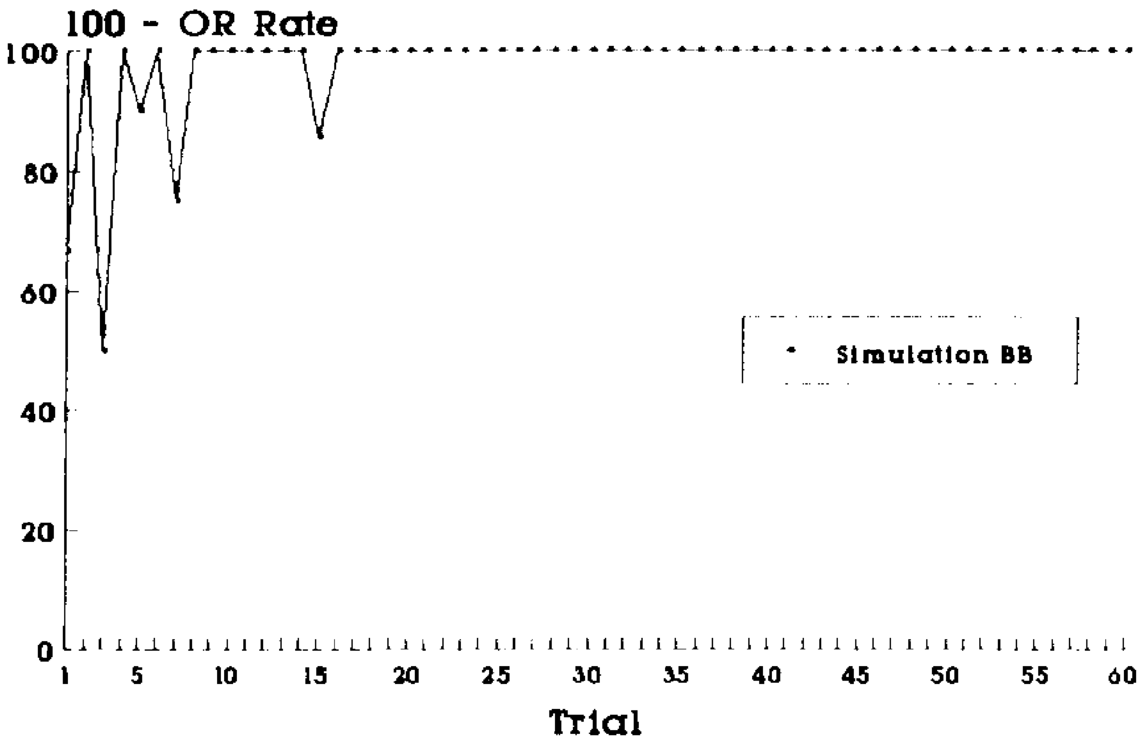


Figure 49. SDM simulations of P&H(70), Simulation BB: No evidence of U-shaped learning.

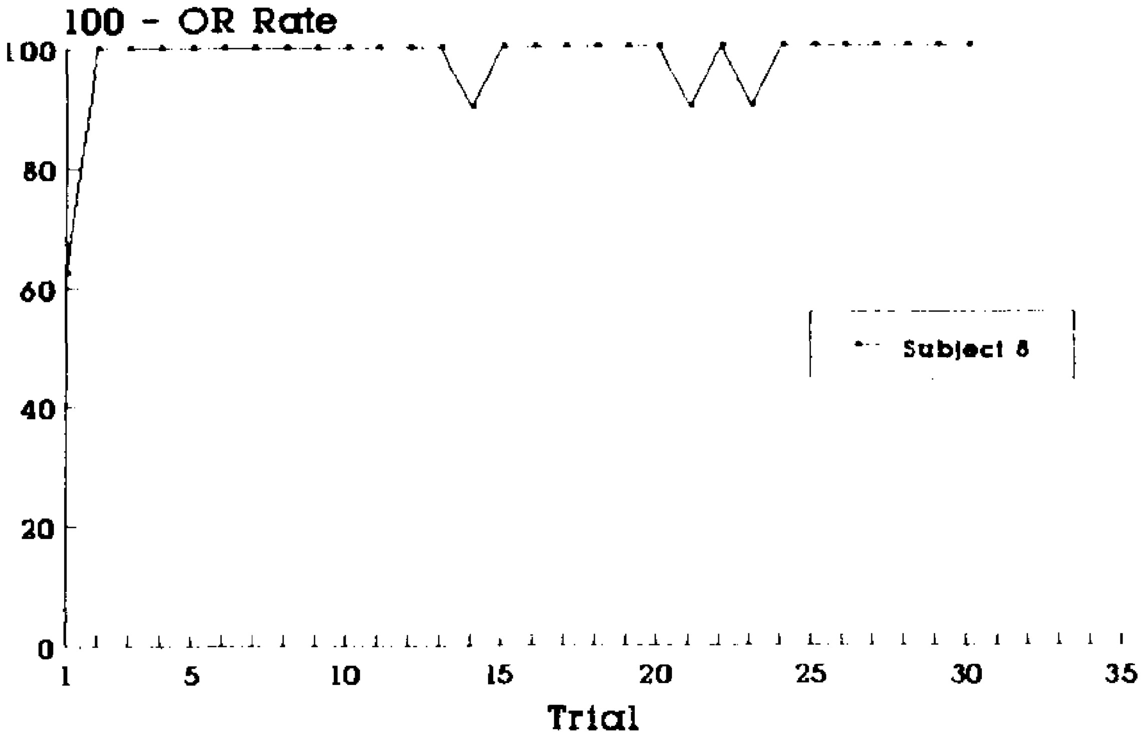


Figure 50. Palermo and Howe (1970), Subject #8: U-shaped learning after an initial period during which there are OR errors.

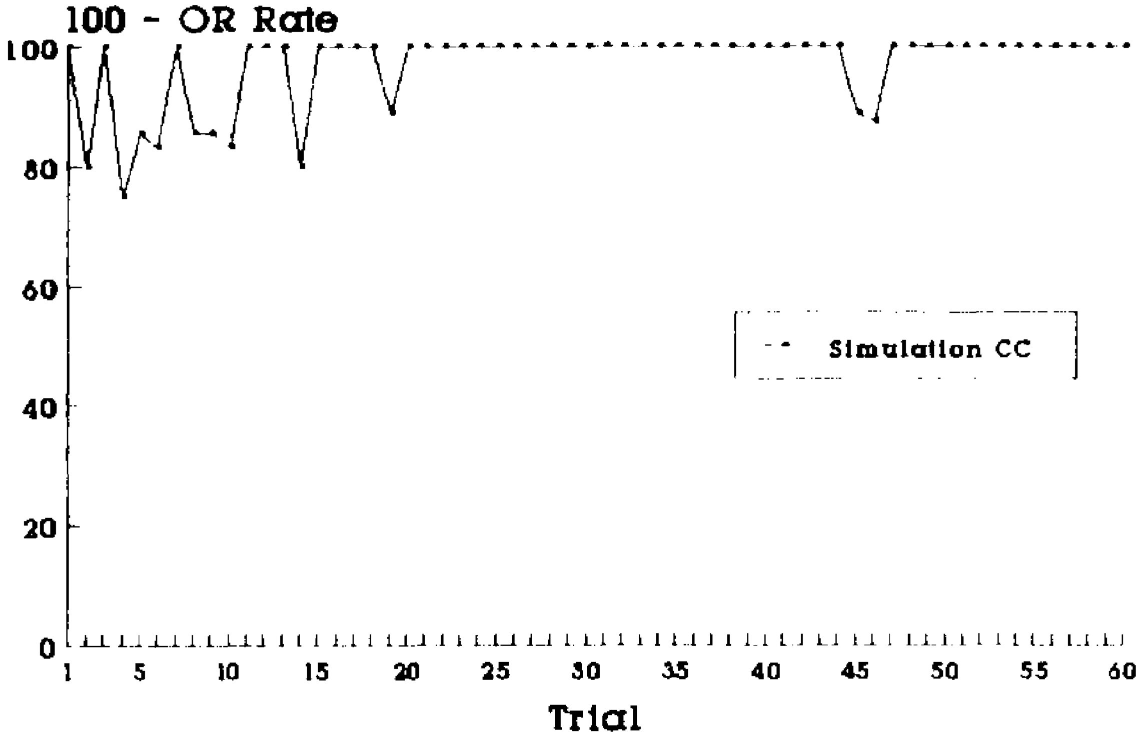


Figure 51. SDM simulations of P&H (70), Simulation CC: U-shaped learning after an initial period during which there are OR errors.

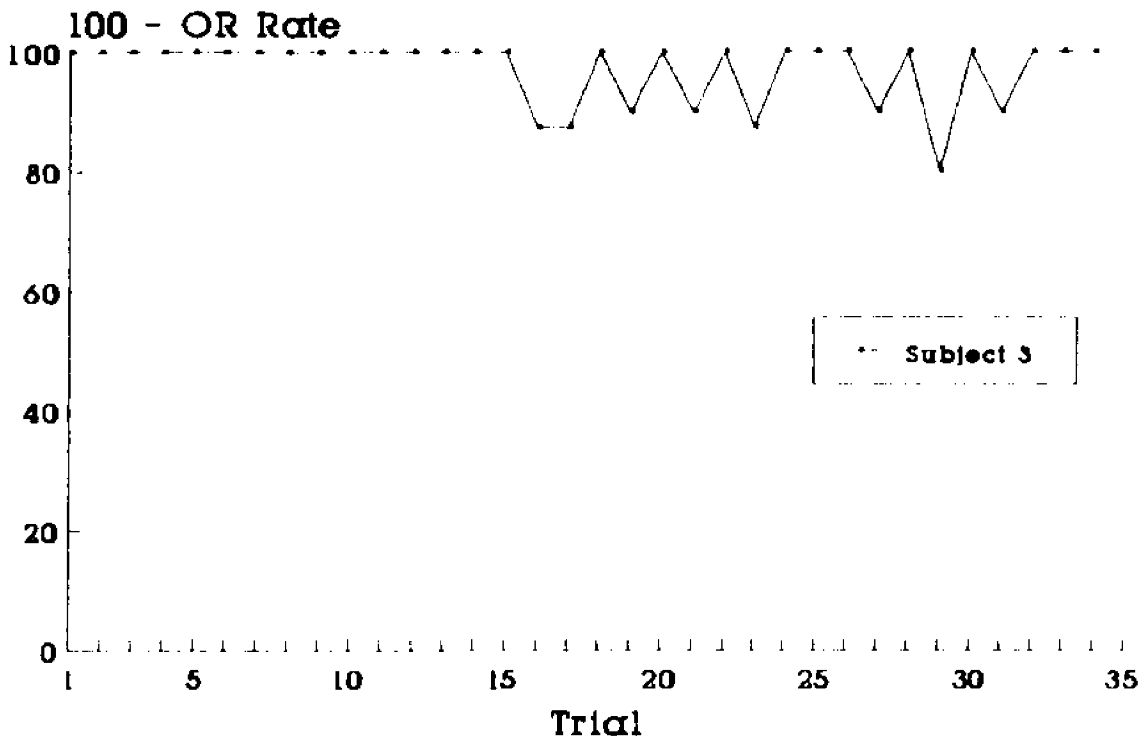


Figure 52. Palermo and Howe (1970), Subject #3: Classic U-shaped learning.

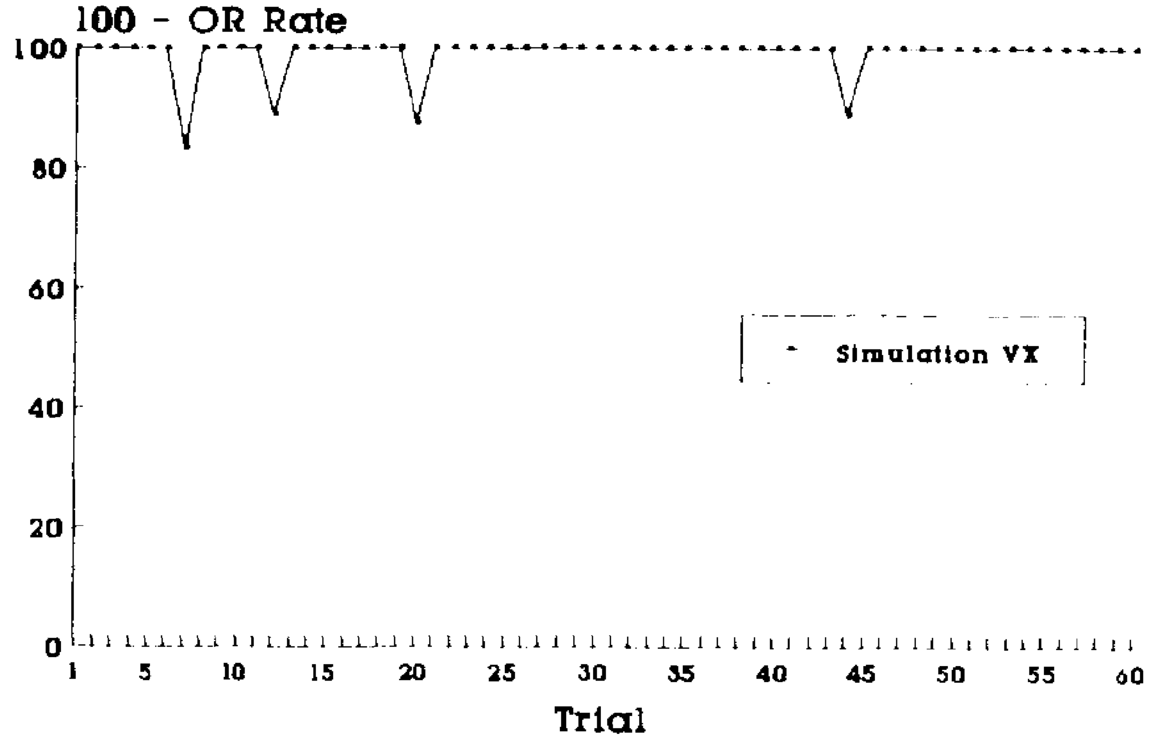


Figure 53. SDM simulations of P&H (70), Simulation VX: Classic U-shaped learning.

These SDM simulations were only unable to produce two of P&H's results. First, a comparison of acquisition curves for the irregulars, based on frequency of presentation, indicated that the acquisition curve for the irregular presented twice per trial was most similar to the acquisition curve for the irregular presented three times for P&H's subjects. In contrast, for the simulations, the acquisition curves for the irregular presented four times and the irregular presented three times most closely paralleled each other. Secondly, the results achieved by P&H's subjects indicated that as subjects became more proficient on the regulars, OR errors accounted for a higher percentage of total errors. In the simulations, on the other hand, OR errors accounted for a decreasing proportion of the total errors as proficiency on the regulars increased. These dissimilarities should not necessarily be taken as evidence that SDM is unable to produce these effects. Instead, as previously discussed (see section II B.4), the problem may lie with the particular coding system used in these simulations. When representing the pairs in these simulations, an attempt was made to make the numbers and letters equidistant from each other. In all probability this was an unwise strategy, since such a strategy reduces the associations between individual letters and numbers. Since humans often make use of associations between letters and numbers, a coding system which takes into account the types of letter associations used by humans (for example, the categorization of letters as vowels and consonants), would have been more consistent with the situation encountered behaviorally. Such a coding system might have produced results more similar to P&H's results.

However, more damaging to the model's ability to mimic human behavior, and a possible reason for the discrepancy between OR rates as regular proficiency increased, is the concept of "guessing". As implemented here, SDM does not guess, it either produces a response, or it returns a string of zeros (which may be considered analogous to a human "omission"). Humans, on the other hand, often guess when they are unsure, and in an experimental situation they may guess more frequently than in a natural setting, since they may feel more obliged to respond. This notion is borne out by the much higher omission rates for simulations, when compared to P&H's subjects. As previously discussed (see section II.B.4), guessing by P&H's subjects may have contributed to the higher OR rates seen as regular proficiency increased. If results more similar to the behavioral evidence are to be realized in SDM simulations, it may be necessary to modify SDM in order to allow it to guess. Coding representations more similar to human representations, and way of handling "guessing", provide two directions for future SDM research.

It should also be noted that the method used in the SDM simulations varied somewhat from P&H's experiment. Some of the pairs in the stimulus-response set were omitted during the study phase in P&H's experiment. These omitted items were later presented to P&H's subjects during the test phase, to see if they could generalize to novel stimuli. In contrast, in the SDM simulations of P&H's experiment, all pairs were presented during each training trial (equivalent to P&H's study phase), and therefore, SDM was never tested on novel items during testing. Since the ability to generalize to novel items is an important human cognitive ability, if SDM is a viable model of human memory, it should be capable of this type of generalization. This issue was not addressed in these SDM simulations of P&H's experiment since preliminary work with SDM (such as the previously described abstraction of the prototypical "X", see Figure 45) strongly suggests that SDM is indeed able to generalize. However, another direction for future research with SDM should certainly address this procedural discrepancy.

Although P&H claim that their data support a two process theory of language acquisition, the success of these SDM simulations challenges this interpretation. Taken along with SDM's successful modelling of the "rule of 78", the results of the P&H simulations lend further support to the view that two separate processes may not be essential in human learning situations which would traditionally be described as containing rules and exceptions. Instead humans may use a single associative strategy, similar to that utilized by SDM. The final evaluation of SDM's ability to simulate specific human cognitive behavior is based on a comparison of the SDM simulations to the F&E experiment. The results of the SDM simulations will also be compared to the connectionist simulations of Jared et al., (1990) results, which were performed by Seidenberg and McClelland (1989) and Daugherty and Seidenberg (1992). In addition, since this experiment offers new insight into human behavior, its results will also be compared to the relevant literature.

One of Jared's et al., major finding was that naming latencies were affected by the composition of a word's neighborhood, with the greatest effect being exerted by the frequency of enemies in the neighborhood. This effect was also in evidence in the F&E experiment, since more errors were observed, during the last 10 trials of the experiment, for words with high frequency enemies than for words with low frequency enemies. The SDM simulations mimicked this result nicely (see Figures 32 and 38 for a comparison of the simulations to the subject data). Although the simulations showed a more significant effect than the F&E subjects, this difference can most probably be attributed to the fact that the simulations showed more complete mastery of the task (as evidenced by perfect performance by the end of the simulations) than the subjects did. Seidenberg and

McClelland's (1989) model of the Jared et al., data also showed a significant difference between the LFF-HFE words and the HFF-LFE words (see Jared et al., 1990, figure 4, p. 698), however their model's effect was not as strong as that observed in the Jared et al., experiment. Of course, Jared et al., also looked at two intermediary friend and enemy conditions (HFF-HFE and LFF-LFE) and found a graded effect based on the frequency of enemies (see Figure 30). These conditions were not examined either the F&E experiment or the SDM simulations. However, it should be pointed out that the Seidenberg and McClelland model was not able to simulate this graded effect. Whether or not SDM is able to simulate this type of graded effect provides yet another area of future research.

Because the traditional two process theory of past tense acquisition suggests that frequency effects should be observed for irregulars, but not for regulars, frequency effects were also addressed in the F&E experiment and simulations. As pointed out by both Daugherty and Seidenberg (1992) and Pinker (1991), the traditional view is supported by behavioral evidence in naming tasks (Prasada, Pinker & Snyder, 1990). However, the F&E experiment provides preliminary evidence that there may be frequency effects for regular, as well as irregular, items since an analysis of the last 10 F&E trials indicates that low frequency pairs show significantly more errors than high frequency pairs. In addition, Prasada, Pinker and Snyder (1990) looked at naming latencies, while the F&E experiment examined error rates. Different measures of performance may not always show similar results, as evidenced by the fact that the reaction time data in the F&E experiment did not show an effect for the frequency of enemies (see section II.C.1.b), while the error rates did.

As expected, since SDM is a one process system, the SDM simulations of F&E indicated that frequency effects were observed for both the regular and irregular pairs. Therefore, SDM simulated the significant frequency effects observed during the last 10 trials of the F&E experiment, but not account for the nonsignificant effect of frequency over all trials. However, since the F&E experiment challenges the view that there are no frequency effects for regular items, the inability of SDM to simulate the results over all F&E trials should not necessarily be considered a severe handicap. Nevertheless, Daugherty and Seidenberg (1992), in their connectionist model, were able to simulate frequency effects for irregular items while showing no frequency effects for regular items. This suggests that one process theories may be able to account for this effect, if further behavioral results support it. Future work with SDM should address its ability to simulate this traditionally supported effect.

Consistency effects were also addressed by the F&E

experiment, since Glushko (1979) and Jared et al., (1990) provided evidence that consistency effects are evident in naming tasks. The traditional two process theory predicts that irregular items should show greater consistency effects (in terms of number of errors and naming latency) than the regulars. This prediction is supported by Glushko, Jared et al., experiment 4, the F&E experiment and the F&E simulations. However, in their naming tasks, Glushko and Jared et al., have provided evidence that not all regular items are generated in the same way. Regular consistent items (which are pronounced by application of the pronunciation rules, and whose neighbors are also all pronounced following these rules) are named faster than regular inconsistent items (which are also pronounced by applying the pronunciation rules, but their neighborhoods include irregularly pronounced words). The traditional two process theory cannot predict this result, since it considers all regulars as equal. Seidenberg (in press) also provides evidence for this phenomenon in past tense acquisition.

Over all trials, the results of the F&E experiment appear to support the traditional two process account, since there was no significant difference in error rates for the regular INC and regular CON pairs. However, an examination of the last 10 trials of the F&E subjects did reveal a difference, which approached significance, between the regular INC and regular CON items. Since the naming experiments performed by Glushko and Jared et al., both test their subjects ability to name items that **have already been mastered**, the effect observed during the last 10 trials of the F&E experiment must be considered. As previously mentioned (see section II.C.3), if the F&E subjects had been run until mastery of the task was evident, the difference between the regular INC and regular CON pairs might have become significant. Although the SDM simulations of F&E did show a significant difference between the regular INC and regular CON items, it cannot be conclusively ruled out that the SDM simulations were unable to simulate the behavioral results, since the simulations were run until the task was mastered, while the F&E subjects were not.

It is unclear whether or not the F&E subjects would have shown the effects demonstrated by Glushko and Jared et al had they been run until they showed evidence of mastery of the task. Therefore, the contention here is that further analysis of the behavioral phenomenon is necessary before any conclusion can be made about SDM's ability to simulate the human data. In addition, should this effect be supported behaviorally, it is clear that one process models are capable of simulating the effect, as evidenced by the results of both these SDM simulations and the simulations performed by Daugherty and Seidenberg (1992).

The final point of analysis of SDM's ability to simulate the F&E experiment deals with U-shaped learning and overregularization. Here the comparisons are striking. Both situations show almost identical U-shaped curves, a substantial decrease in performance on the irregulars immediately following irregular criterion, and evidence of a return to perfect irregular performance. Taken along with Palermo and Howe's (1970) results, the fact that the F&E experiment showed U-shaped learning, accompanied by OR, lends further support to the idea that these effects may be considered general cognitive phenomena, rather than phenomena limited to the domain language acquisition. In addition, SDM's ability to simulate these phenomena, in yet another experimental situation, also furnishes further evidence that U-shape learning and OR need not be explained by a traditional two process system. Therefore, the results of all three sets of SDM simulations strongly suggest that a one process associative memory system is able to account for much of the behavioral data. These results suggest that SDM may in fact be a viable model of human memory.

A number of references have already been made to various avenues of future research with SDM. However, one notable area remains to be addressed. Based on the fact that the success of all these SDM simulations relied on a strategy of decreasing noise, Martin Chodorow (personal communication, April 16, 1993) has developed a modification of Kanerva's (1988) original SDM model, similar in some respects to suggestions made by Keeler (1988). As described previously (see the introduction to section II and Figure 5), when noise is added to an address in SDM, in effect the storage area around the original address is increased. Chodorow's modification of SDM takes unused hard memory locations and places them into areas of memory which are being addressed. In effect, this makes the used areas of memory more dense, and the unused areas of memory more sparsely populated. This, in turn, allows greater discrimination in the areas of greatest interest, that is, the areas of memory which are being addressed. Preliminary work indicates that in this modified SDM model, reducing the storage and retrieval radii produces results very similar to the results achieved by the decreasing noise strategy used in this work. Future research will examine the ability of this modified SDM model to simulate human behavior, in the hopes of further supporting the view that SDM is a viable model of human memory.

APPENDIX A

Individual SDM simulations of the "rule of 78"

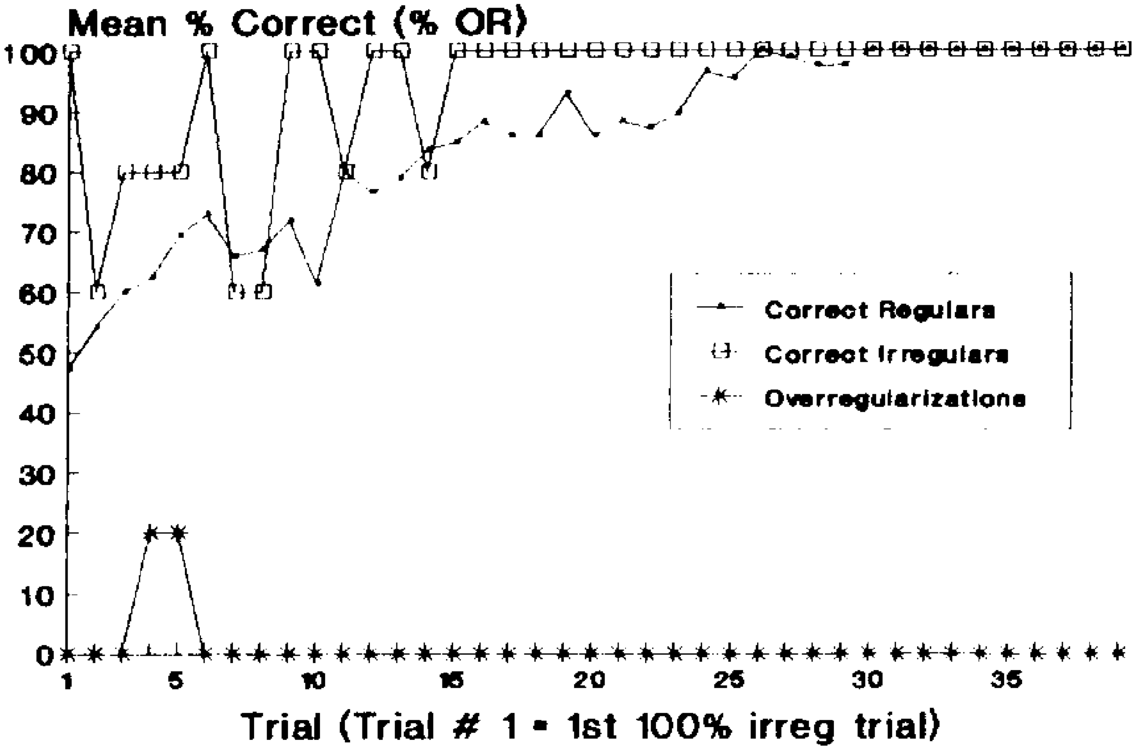


Figure AP-A1. Mean results by trial for SDM simulation DX (one of the 10 simulations of the rule of 78), starting at the trial in which simulation DX first reached 100% correct on the irregular. Overregularization errors (%OR) apply only to the irregular.

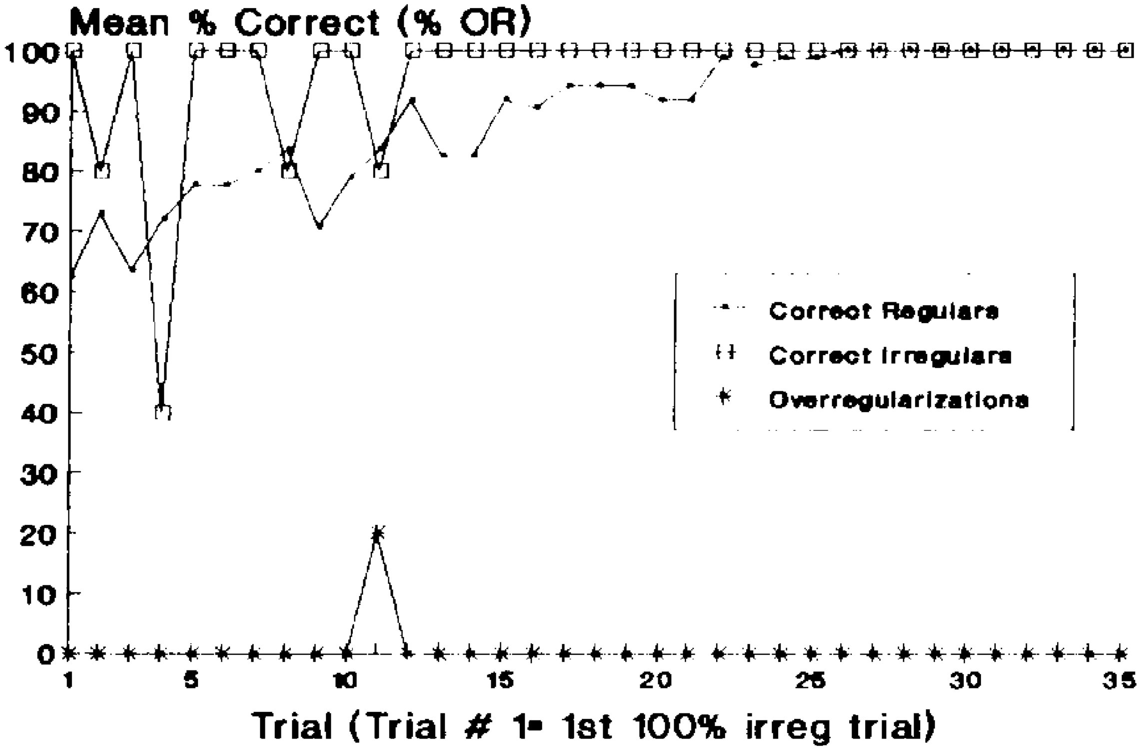


Figure AP-A2. Mean results by trial for SDM simulation PX (one of the 10 simulations of the rule of 78), starting at the trial in which simulation PX first reached 100% correct on the irregular. Overregularization errors (%OR) apply only to the irregular.

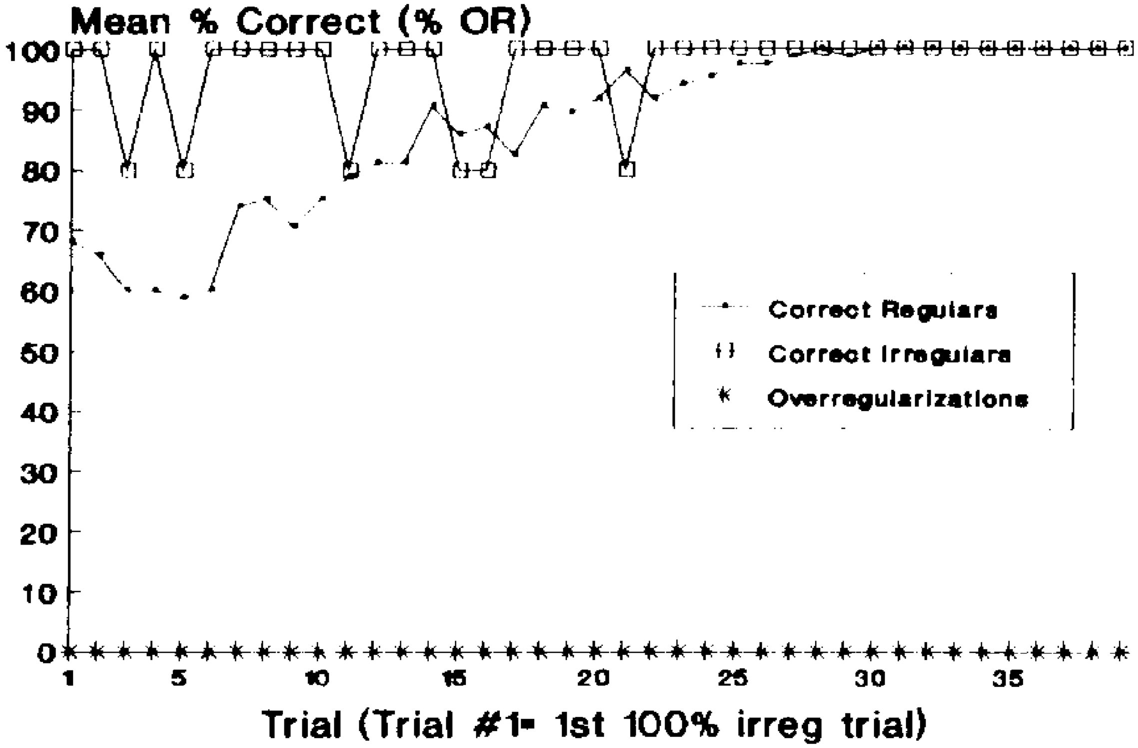


Figure AP-A3. Mean results by trial for SDM simulation QX (one of the 10 simulations of the rule of 78), starting at the trial in which simulation QX first reached 100% correct on the irregular. Overregularization errors (%OR) apply only to the irregular.

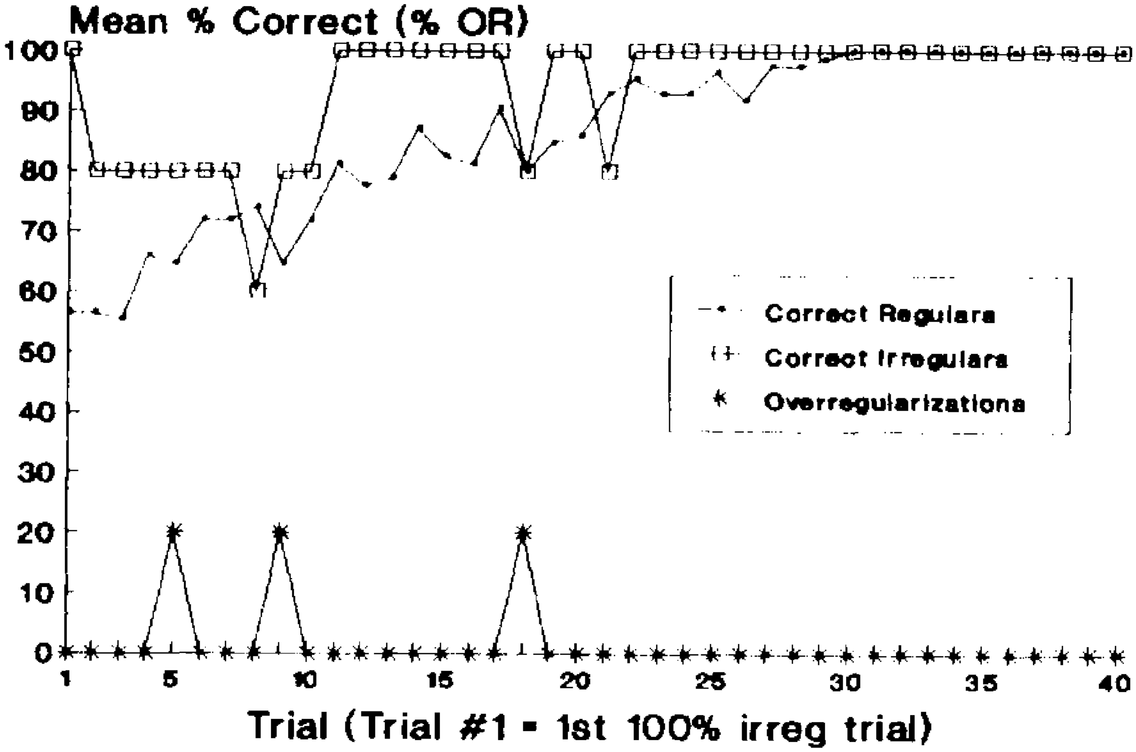


Figure AP-A4. Mean results by trial for SDM simulation RX (one of the 10 simulations of the rule of 78), starting at the trial in which simulation RX first reached 100% correct on the irregular. Overregularization errors (%OR) apply only to the irregular.

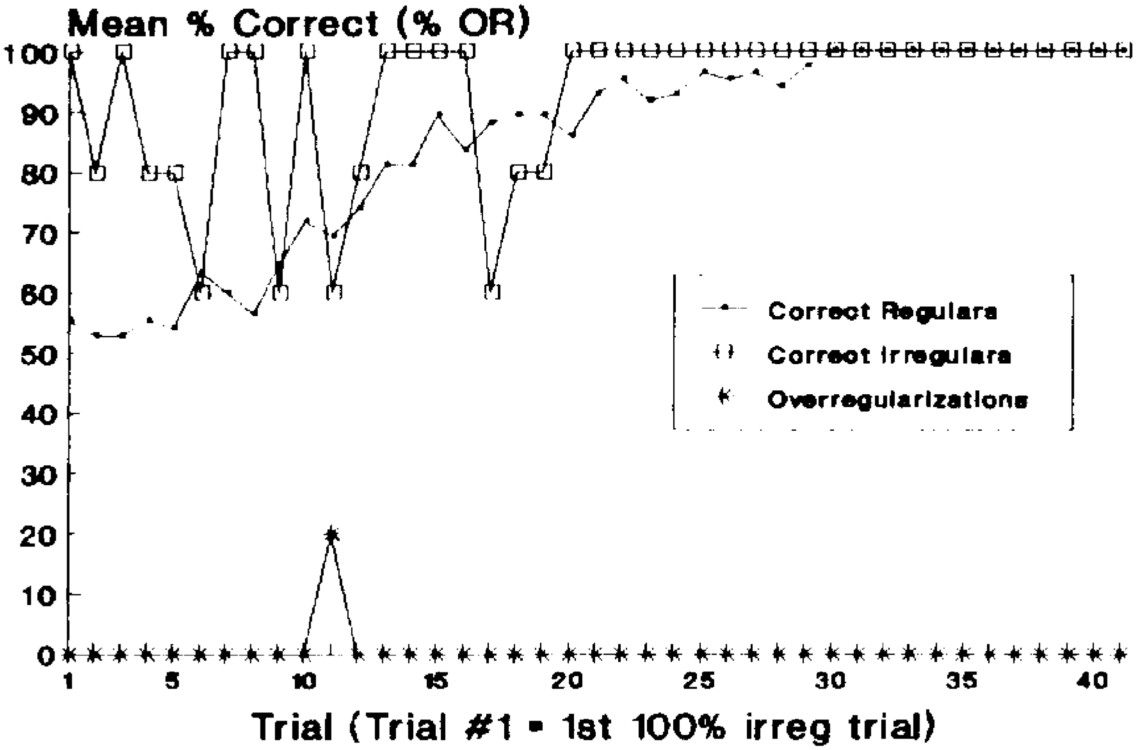


Figure AP-A5. Mean results by trial for SDM simulation TX (one of the 10 simulations of the rule of 78), starting at the trial in which simulation TX first reached 100% correct on the irregular. Overregularization errors (%OR) apply only to the irregular.

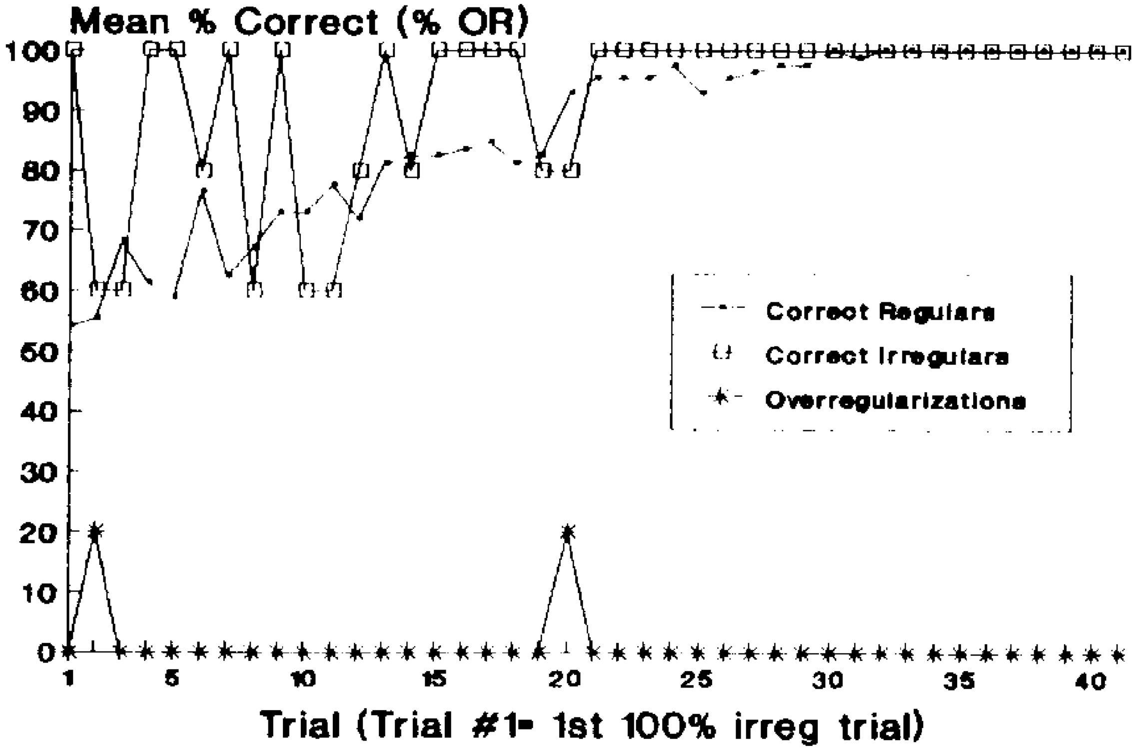


Figure AP-A6. Mean results by trial for SDM simulation UX (one of the 10 simulations of the rule of 78), starting at the trial in which simulation UX first reached 100% correct on the irregular. Overregularization errors (%OR) apply only to the irregular.

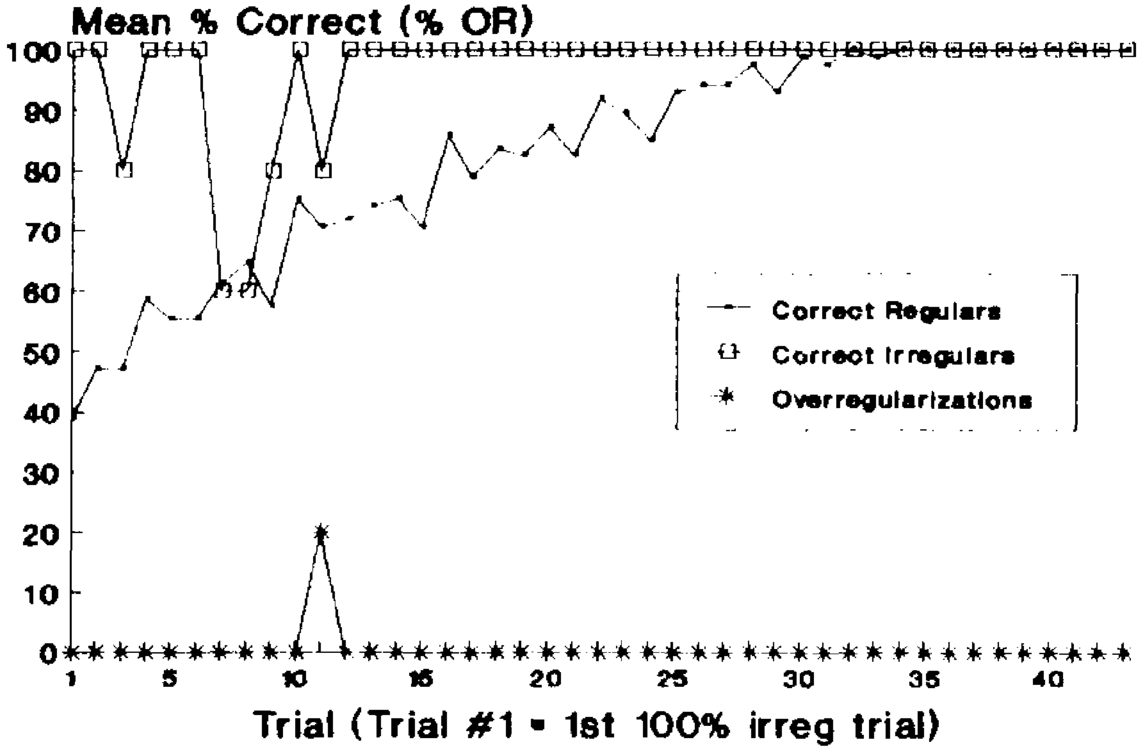


Figure AP-A7. Mean results by trial for SDM simulation VX (one of the 10 simulations of the rule of 78), starting at the trial in which simulation VX first reached 100% correct on the irregular. Overregularization errors (%OR) apply only to the irregular.

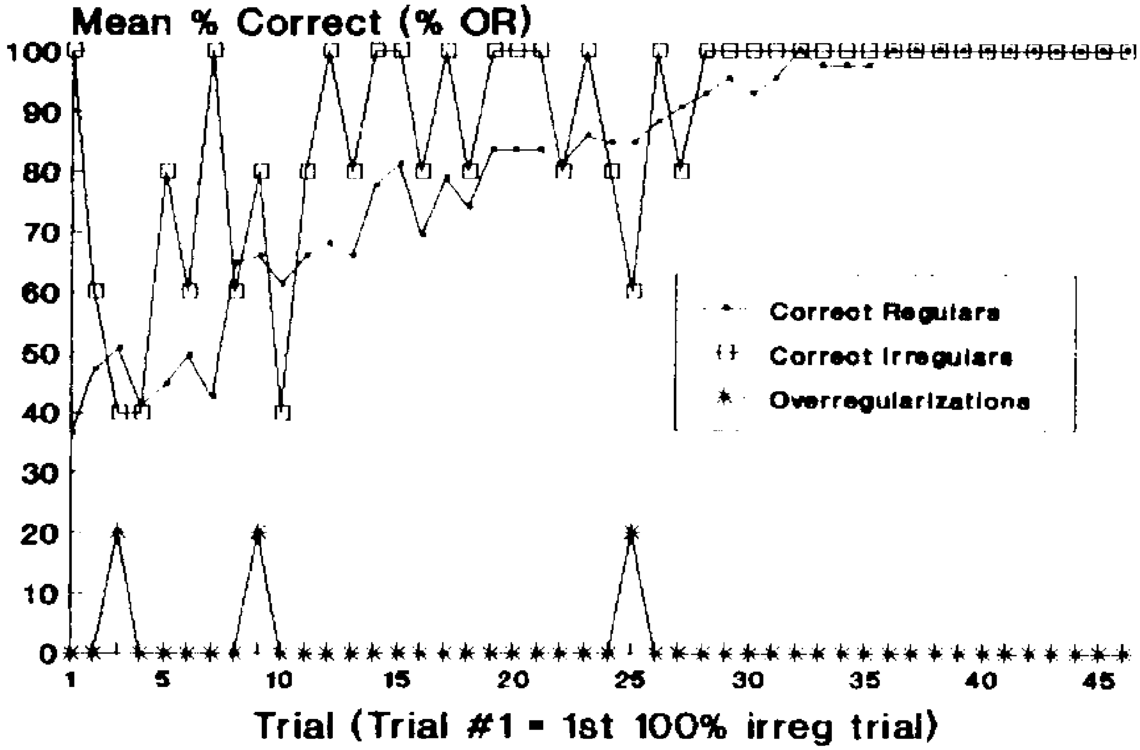


Figure AP-A8. Mean results by trial for SDM simulation WX (one of the 10 simulations of the rule of 78), starting at the trial in which simulation WX first reached 100% correct on the irregular. Overregularization errors (%OR) apply only to the irregular.

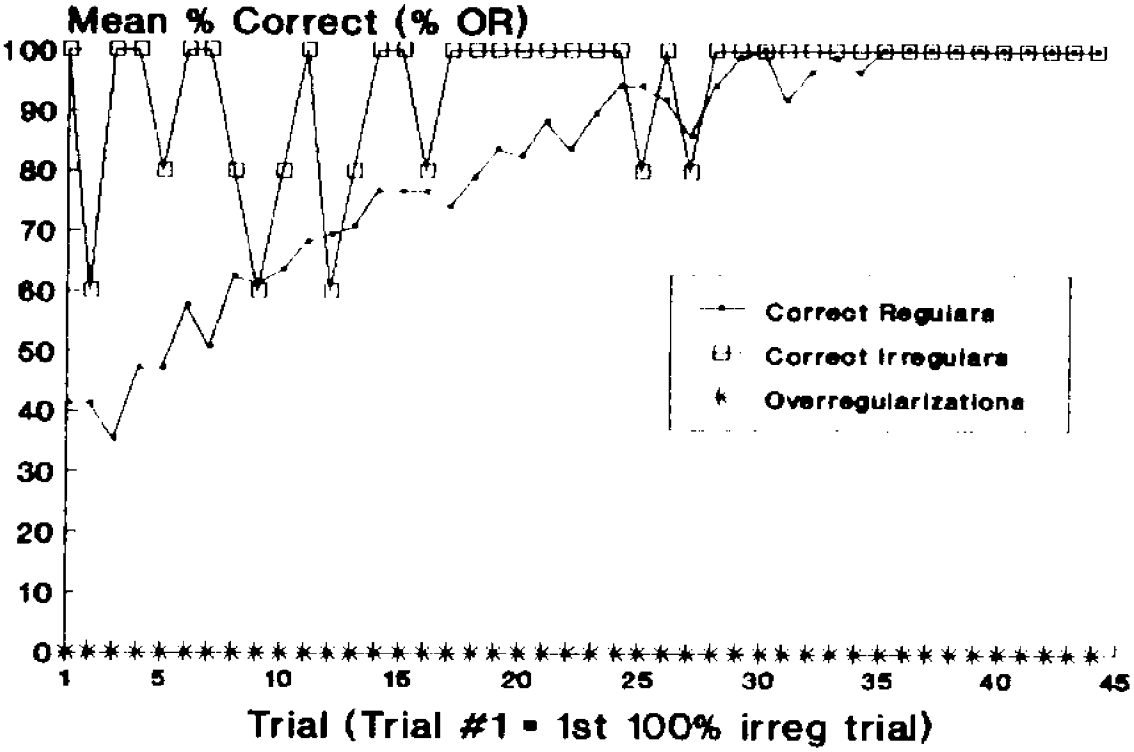


Figure AP-A9. Mean results by trial for SDM simulation XX (one of the 10 simulations of the rule of 78), starting at the trial in which simulation XX first reached 100% correct on the irregular. Overregularization errors (%OR) apply only to the irregular.

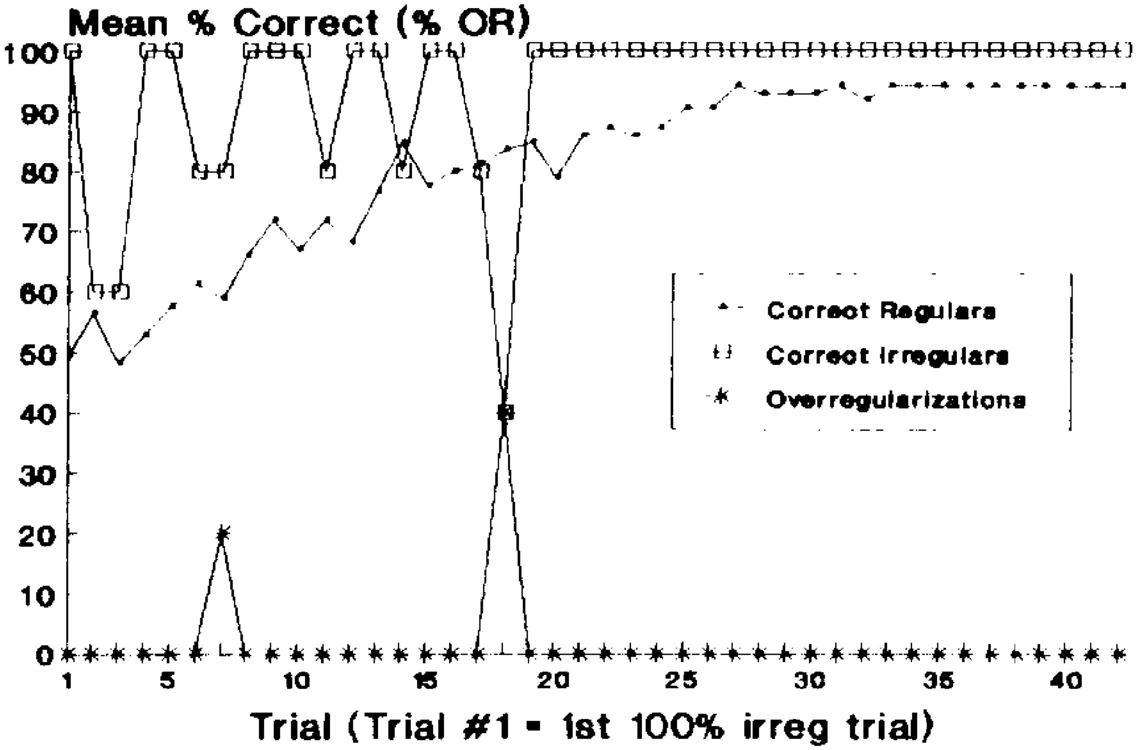


Figure AP-A10. Mean results by trial for SDM simulation YX (one of the 10 simulations of the rule of 78), starting at the trial in which simulation YX first reached 100% correct on the irregular. Overregularization errors (%OR) apply only to the irregular.

Appendix B

Instructions to subjects in F&E Experiment

"This is an experiment to determine how quickly you can learn to associate pairs of three digit numbers.

Part I of the experiment is a study period. A three digit number (the STIMULUS) will be displayed in the center of the computer screen. After a few seconds, the three digit number RESPONSE associated with that stimulus number will be displayed below it. This pair will remain on the screen for a few seconds. PLEASE STUDY IT CAREFULLY! There are 13 different STIMULUS-RESPONSE pairs in the list you have to learn.

PART II of the experiment is the learning phase. During this phase, one of the three digit STIMULUS numbers will appear on the screen for a few seconds. Your task is to type in the three digit RESPONSE associated with that stimulus AS QUICKLY AS POSSIBLE. Use the number pad, on the right hand side of the keyboard, for your response. You will have 7 seconds to respond. Your response will be displayed in the lower left hand corner of the computer screen and then the correct response will be displayed in the center of the screen, under the stimulus number associated with it. Don't worry if you do not respond in the allotted time, just go on to the next stimulus. After all the stimuli are presented, in random order, "End of

Trial" will be displayed on the screen. After a few seconds, the next learning trial will begin.

If you have any questions, please ask them now."

BIBLIOGRAPHY

Berko, J. (1958). The child's learning of English morphology. Word, 14, 150-177.

Brown, R. (1973). A first language: The early stages. Cambridge, MA: Harvard University Press.

Brown, R., Cazden, C. B., & Bellugi, U. (1968). The child's grammar from I to III. In J. P. Hill (Ed.), Minnesota symposium on child development (pp. 28-73). Minneapolis, MN: University of Minnesota Press.

Bybee, J. L., & Slobin, D. I. (1982). Rules and schemas in the development and use of the English past tense. Language, 58(2), 265-289.

Cazden, C. B. (1968). The acquisition of noun and verb inflections. Child Development, 39, 433-438.

Conrad, R. (1963). Acoustic confusions and memory span for words. Nature, 197, 1029-1030.

Conrad, R. (1964). Acoustic confusions in immediate memory. British Journal of Psychology, 55, 75-84.

Daugherty, K., & Seidenberg, M. S. (1992). Rules or connections? The past tense revisited. Proceedings of the Fourteen Annual Meeting of the Cognitive Science Society, Hillsdale, NJ: Erlbaum.

Denning, P. J. (1989). The science of computing: Sparse distributed memory. American Scientist, 77, 333-335.

Denning, P. J. (1992). The science of computing: Neural networks. American Scientist, 80, 426-429.

Ervin, S.M. (1964). Imitation and structural change in children's language. In E. H. Lenneberg (Ed.), New directions in the study of language (pp. 163-189). Cambridge, MA: MIT Press.

Ervin-Tripp, S. M. (1971). An overview of theories of grammatical development. In D. I. Slobin (Ed.), The ontogenesis of grammar: A theoretical symposium (pp. 189-212). New York: Academic Press.

Esper, E. A. (1925). A technique for the experimental investigation of associative interference in artificial linguistic material. Language Monographs, 1, 1-47.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. Cognition, 28, 3-71.

Francis, W. N., & Kucera, H. (1982). Frequency analysis of English usage: Lexicon and grammar. Boston, MA: Houghton Mifflin Company.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. Journal of Experimental Psychology: Human Perception and Performance, 5, 674-691.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. Journal of Memory and Language, 29, 687-715.

Kanerva, P. (1988). Sparse distributed memory. Cambridge, MA: Bradford Books / MIT Press.

Karmiloff-Smith, A., & Inhelder, B. (1974/75). If you want to get ahead, get a theory. Cognition, 3(3), 195-212.

Keeler, J. D. (1988). Comparison between Kanerva's sparse distributed memory and Hopfield-type neural networks. Cognitive Science, 12, 299-329.

Kucera, H., & Francis, N. (1967). Computational analysis of present-day American English. Providence, RI: Brown University Press.

Kuczaj, S. A. (1977). The acquisition of regular and irregular past tense forms. Journal of Verbal Learning and Verbal Behavior, 16, 589-600.

Kuczaj, S. A. (1978). Children's judgements of grammatical and ungrammatical irregular past-tense verbs. Child Development, 49, 319-326.

Lachter, J., & Bever, T. G. (1988). The relation between linguistic structure and associative theories of language learning- A constructive critique of some connectionist learning models. Cognition, 28, 195-247.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. Cognition, 40, 121-157.

Marchman, V. A., & Bates, E. (in press). Continuity in lexical and morphological development: A test of the critical mass hypothesis. Journal of Child Language.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, *57*,(4, Serial No. 228).

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation: Vol. 24. Advances in research and theory* (pp. 109 - 165). San Diego, CA: Academic Press.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

Palermo, D. S., & Eberhart, V. L. (1968). On the learning of morphological rules: An experimental analogy. *Journal of Verbal Learning and Verbal Behavior*, *7*, 337-344.

Palermo, D. S., & Howe, H. E. (1970). An experimental analogy to the learning of past tense inflection rules. *Journal of Verbal Learning and Verbal Behavior*, *9*, 410-416.

Pinker, S. (1991). Rules of language. *Science*, *253*, 530-535.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*, 73-193.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, *38*, 43-102.

Plunkett, K., & Marchman, V. (in press). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*.

Prince, A., & Pinker, S. (1988). Rules and connections in human language. *Trends in Neuroscience*, *11*(5), 195-202.

Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review*, *97*(2), 285-308.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland & D. E, Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2 Psychological and biological models* (pp. 216-271). Cambridge, MA: Bradford Books / MIT Press.

Seidenberg, M. S. (in press). A connectionist modeling approach to word recognition and dyslexia. *Psychological Science*.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523-568.

Slobin, D. I. (1966). The acquisition of Russian as a native language. In F. Smith & G. A. Miller (Eds.), *The genesis of language: A psycho-linguistic approach* (pp. 129-148). Cambridge, MA: MIT Press.

Slobin, D. I. (1971a). Data for the symposium. In D. I. Slobin (Ed.), *The ontogenesis of grammar: A theoretical symposium* (pp. 3-14). New York: Academic Press.

Slobin, D. I. (1971b). On the learning of morphological rules: A reply to Palermo and Eberhart. In D. I. Slobin (Ed.), *The ontogenesis of grammar: A theoretical symposium* (pp. 215-223). New York: Academic Press.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. Ferguson & D. I. Slobin (Eds.), *Studies of child language development* (pp 175-208). New York: Holt Rinehart and Winston.

Strauss, S. (Ed.). (1982). *U-shaped behavioral growth*. New York: Academic Press.

Wickelgren, W.A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior.*Psychological Review*, *76*, 1-15.