Implementations are not Conceptualizations: Revising the Verb Learning Model

Brian MacWhinney
Carnegie Mellon University, macw@cmu.edu

Jared Leinbach
Carnegie Mellon University

Follow this and additional works at: http://repository.cmu.edu/psychology
Implementations are not Conceptualizations:
Revising the Verb Learning Model

Brian MacWhinney
Jared Leinbach
Carnegie Mellon University

Address for correspondence:

Brian MacWhinney
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
ABSTRACT

In a recent issue of this Journal, Pinker and Prince (1988) and Lachter and Bever (1988) presented detailed critiques of Rumelhart and McClelland's (1986) connectionist model of the child's learning of the phonological form of the English past tense. In order to address these criticisms, a new connectionist model was constructed using the back-propagation algorithm, a larger input corpus, a fuller paradigm, and a new phonological representation. This new implementation successfully addressed the criticisms of the phonological representation used by Rumelhart and McClelland. It did a much better job of learning the past tense using a fuller input set with realistic frequencies of occurrence. Ancillary simulations using the same network were able to deal with the homonymy problem and the generation of forms like “ated” from “ate.” The one feature not provided by the new model was a way of modeling early correct production of irregular forms. The success of the new model can be used to help clarify the extent to which the published critiques apply to a particular connectionist implementation as opposed to fundamental principles underlying the broader connectionist conceptualization.

---

1 This work was supported by an NICHD grant to Brian MacWhinney and an ONR grant to Jay McClelland. Our thanks to Jay McClelland, Elizabeth Bates, Kim Plunkett and Virginia Marchman for comments on this paper. Jay McClelland also helped set up the English input set and configure the simulation.
Connectionism is "in." Not since the Dark Ages of the pre-Chomskyan era have we seen so much interest in associationist models of human thinking. Streaming forth from their banishment in the Skinnerian dungeons are dozens of detailed computational models based on the new language of networks, nodes, and connections. Riding on the crest of this new wave, the Parallel Distributed Processing (PDP) models of Rumelhart, McClelland, and their colleagues stand out as some of the most interesting applications of connectionism to substantive problems in cognitive science. Proponents of these new models present them as a major challenge to the ancien régime -- a definitive revolution in the way in which we understand the human mind. Yet, as we all know, revolutions in academia are seldom bloodless. Inevitably, the proponents of the new paradigm tend to overstate their case and, inevitably, the Old Guard tries to form a unified front to challenge the contributions of the newcomers. The ensuing confusion infuriates some, galvanizes others, and perplexes everyone. All of these things are happening now in the Great Debate that is taking place between the New Connectionism and the Classical Model.

A recent issue of this journal (Pinker & Mehler, 1988) was devoted to an in-depth criticism of the PDP agenda. At the focus of much of this criticism was a particular PDP simulation by Rumelhart and McClelland (1986; 1987) of the acquisition of the phonological form of the English past tense. A lengthy article by Pinker and Prince (1988) was devoted entirely to an in-depth critique of the R&M verb learning model. A somewhat shorter article by Lachter and Bever (1988) devoted about half of its pages to the R&M verb learning model, while also critiquing four other connectionist models in somewhat less detail. In all, more than 140 journal pages were devoted to the critique of the verb learning model -- nearly three times the number of printed pages devoted to the original article.

The chapter that served as the target of this extensive criticism had as its goal the formulation of a computationally explicit connectionist model of the child's learning of the past tense forms of English verbs. This "verb-learning" model of Rumelhart and McClelland (henceforth R&M) was capable of taking as input a present tense verb, such as "ring," and providing as output a past tense form, such as "rang." It did this without any overt encoding of a set of rules and without any formal construction of morphological paradigms.

“We have, we believe, provided a distinct alternative to the view that children learn the rules of English past-tense formation in any explicit sense. We have shown that a reasonable account of the acquisition of past tense can be provided without recourse to the notion of a 'rule' as anything more than a description of the language. We have shown that, for this case, there is no induction problem. The child need not decide whether a verb is regular or irregular. There is no question as to whether the inflected form should be stored directly in the lexicon or derived from more general principles. There is not even a question (as far as generating the past-tense form is concerned) as to whether a verb form is one encountered many times or one that is being generated for the first time. A uniform procedure is applied for producing the past-tense form in every case.”
(Rumelhart & McClelland, 1987, p. 246)

There is obviously a great deal at stake here. On the one hand, rules and rule-like symbolic structures are fundamental to many influential current approaches in linguistics, philosophy, and psychology. A great deal of work has been done in the context of rule systems. Questioning this approach would also call into question an enormous body of linguistic and psycholinguistic research and could force researchers to undergo a major conceptual reorganization. On the other hand, it could be that it is the connectionist account that is flawed. The R&M model is not the only rule-less, cue-based, account that
has been developed. There are now cue-based accounts for the acquisition of German morphology (MacWhinney, Leinbach, Taraban & McDonald, 1989; Taraban, McDonald & MacWhinney, 1989), Hungarian and Turkish morphology (Hare, 1990), phonological rules generally (Touretzky & Wheeler, 1989), prosodic structure (Gupta & Touretzky, 1994), semantic structure (MacWhinney, 1989; 1990), and syntactic structure (Elman, 1990; McClelland & Kawamoto, 1986). If these new cue-based connectionist models are fundamentally flawed, we need to understand this before too much work would be wasted going in the wrong direction.

Sensing the seriousness of these issues, Pinker and Prince (henceforth P&P) and Lachter and Bever (henceforth L&B) constructed their analyses carefully and thoughtfully. Their critique has generated a great deal of useful discussion. Moreover, as will become clear in the next pages, there is good reason to agree with many of the detailed aspects of their analyses. Much to their credit, P&P and L&B focused their critique not simply on the overall connectionist conceptualization; rather they targeted in on the detailed claims and predictions of the particular models which implement the overall conceptualization. This is a respected and time-honored strategy in scientific discourse. Indeed, it is hard to imagine how science could progress in any other way. The logic is as follows. Each of a set of competing conceptualizations is viewed as spawning a set of implementations for particular empirical problems. One then lines up each of the candidate implementations and subjects them to comparative criticism in terms of logical consistency, match to known data, and extendability to further problems. When a winner is declared, the conceptualization underlying the implementation gains some credibility and the alternative conceptualizations are weakened.

There is a further convention which stipulates that, if one has successfully cast some doubt upon the specifics of an implementation, it is legitimate to then generalize from this to a critique of the underlying conceptualization. Both P&P and L&B availed themselves of this opportunity. P&P state their conclusions in the following form.

“We will conclude that the claim that parallel distributed processing networks can eliminate the need for rules and for rule induction mechanisms in the explanation of human language is unwarranted. In particular, we argue that the shortcomings are in many cases due to central features of connectionist ideology and irremediable; or if remediable, only by copying tenets of the maligned symbolic theory. The implications for the promise of connectionism in explicating language are, we think, profound.” (P&P, p. 82)

L&B also believed that they had isolated a fundamental weakness in current connectionist approaches to language learning.

“The connectionist models we have considered arrive at rule-like regularities in language behavior only insofar as the models already contain architectures and devices explained in humans by mental representations of categorical rules.” (L&B, p. 243)

In the bodies of their chapters, both P&P and L&B focus their critique on details of the R&M implementation. However, at the ends of their chapters, they use that critique as a basis for making more general claims about the “shortcomings” of connectionism. This form of argumentation seems entirely fair and proper.

There is another standard for argumentation against which these two critiques fair less well. Both P&P and L&B recognize the importance of fully simulated models as ways of testing theoretical claims. However, neither P&P nor L&B present a fully implemented symbolic alternative to the R&M model. Thus, the competition we are asked to judge is one between connectionist implementations and symbolic principles. As P&P (p. 181) correctly note, “there is a gap between revolutionary manifestos and actual accomplishments.” The construction of a full symbolic model of the acquisition of the
English past tense would do much to fill this gap. With equally detailed candidate implementations from the two competing conceptualizations, we could more properly use implementations as a way of judging the usefulness of underlying conceptualizations. Until that becomes possible, we will need to focus our attention on ways of improving the connectionist alternative, since that is the only detailed model currently in the field.

In this paper we present a new connectionist implementation which we believe can stand successfully against the various criticisms directed by P&P and L&B against the original verb learning model. In order to explore this issue, we will first summarize the original R&M model and the individual criticisms leveled against it. Then we will examine a new connectionist model of verb learning designed to address each of the detailed criticisms of the R&M verb learning model.

1. The Verb Learning Model

The verb learning model made use of a simple two-layer connectionist network which took as input the phonological form of the present tense of an English verb and attempted to produce as output the phonological form of the past tense. In other words, given the form “take,” the model was supposed to output the form “took.” In order to stimulate learning, the network was given a series of learning trials on which it tried to produce the correct form. If it was successful, no change was made in the weights of the connections in the network. If the output was incorrect, the weights were adjusted by a learning rule that moved the network closer to correct performance. On every learning trial, the network was given feedback about its deviations from the correct target.

The input features fed to the network were a series of “Wickelfeatures” or featural decompositions of context-sensitive allophones (Wickelgren, 1969). For example, the segment /b/ in the word “bet” was coded segmentally as a front, voiced, interrupted stop; the segment /E/ was coded as a low, short, front vowel; the segment /t/ was coded as as an unvoiced, middle, interrupted stop; and the beginning and end of the word were coded as “ends.” These segmental codings were then joined into triplets composed of one of the features of the target segment, one of the features of the previous segment, and one of the features of the following segment. For example, some of the triplets representing the word “bet” would be “end -- interrupted -- vowel,” “stop -- low -- stop,” or “voiced -- low -- unvoiced.” Of the 1210 possible triplets of these features, R&M selected 260. This was done to reduce the computational demands of the model. The set of Wickelfeatures describing an input was generated by an encoding network from a stem phoneme string, and a binding network was used to relate the output Wickelfeatures to actual past tense forms in English. The overall architecture of the model is summarized in Figure 1.
In this architecture every input unit was connected to every output unit. Output units were
turned “on” in a probabilistic manner. First a net input to each output unit was determined.
This consisted of the sum of every active input unit times its weight to the output unit, plus
a bias term associated with the output unit. The probability of the output unit actually
coming on was determined using the following logistic function.

\[ p(a_i = 1) = \frac{1}{1 + e^{-\text{net input}/k}} \]

On each trial, the output form was represented by the set of output Wickelfeatures that were
“on.”

Training was done using the classic perceptron convergence procedure (Rosenblatt, 1959).
If a unit should have fired on a given trial, but failed to do so, then the weights to that unit
from active inputs and the unit’s bias were all increased by a small constant amount \( h \).
If a unit fired when it should not have, then the weights from active inputs and its bias were
decreased by \( \eta \).

By making successive adjustments to the biases and the weights on the connections
between the input and output Wickelfeatures, the model learned to correctly produce the
past tense of most English verbs and succeeded in producing correct past tense forms for
verbs it had never seen. R&M claimed that detailed analyses of the course of learning in
the network revealed many parallels between the changes in the model’s behavior over time
and developmental patterns found in English-speaking children.
2. The Criticisms

P&P and L&B felt that claims made by R&M regarding the success of the verb learning model were extremely misleading. They expressed their respective cases against the verb learning model in terms of a set of points.

2.1. Problems raised by Pinker and Prince.

P&P (p. 81) presented at least ten specific problems which they believed were incorrectly addressed by the R&M model. These problems are repeated below in quotation marks. Numbers and problem names have been added for ease of reference. Section numbers are given for the reader who wishes to check the detailed discussion of each of the problems.

1. **The u-shaped learning problem:** “Rumelhart and McClelland’s actual explanation of children’s stages of regularization of the past tense morpheme is demonstrably incorrect.” See section 7.2.1 in P&P and section 5.2 in L&B.
2. **The “ated” problem:** “Their explanation for one striking type of childhood speech error is also incorrect.” See section 7.2.4 in P&P.²
3. **The “hit-hit” problem:** “Their other apparent successes in accounting for developmental phenomena either have nothing to do with the model’s parallel distributed processing architecture, and can easily be duplicated by symbolic models, or involve major confounds and hence do not provide clear support for the model.” See section 7.2.2 (particularly p. 150) in P&P.
4. **The “algalgal” problem:** “The model is incapable of representing certain kinds of words.” See section 4.1 (p. 97) in P&P and section 5.1 in L&B.
5. **The “slit-silt” problem:** “It is incapable of explaining patterns of psychological similarity among words.” See section 4.1 (p. 98) in P&P and sections 3 and 5 in L&B.
6. **The “brag-grab” problem:** “It easily models many kinds of rules that are not found in any human language.” See section 4.1 (p. 100) in P&P.
7. **The phonological regularities problem:** “It fails to capture central generalizations about English sound patterns. It makes false predictions about derivational morphology, compounding and novel words.” See section 4.2 in P&P.
8. **The homophony problem:** “It cannot handle the elementary problem of homophony.” See section 4.3 in P&P and section 3 in L&B.
9. **The convergence problem:** “It makes errors in computing the past tense forms of a large percentage of the words it is tested on.” See section 5 in P&P.
10. **The regular pattern problem:** “It makes incorrect predictions about the reality of the distinction between regular rules and exceptions in children and in languages.” See section 4.4 in P&P and section 3 in L&B.

2.2. Problems raised by Lachter and Bever.

² Under this same point, P&P (see section 7.2.3) also discuss the extent to which the R&M simulation succeeded in generating errors in six different subclasses of irregular verbs (feel-felt, seek-sought, bite-bit, sing-sang, break-broke, and blow-blew) that matched the proportions of these errors in the experimental data of Bybee and Slobin (1982). However, the experimental data of Bybee and Slobin give us only a single time slice from the larger developmental picture and it is not reasonable to attempt to evaluate a full learning model such as that of R&M against a single slice of acquisitional data. Because there are no developmental data on this topic we will not deal with it in our discussions or simulations.
Problems 1, 4, 5, 7, 8, and 10 were also raised by L&B. In addition, L&B constructed a second form of criticism that was quite different from that developed by P&P. L&B claimed that the R&M verb learning model achieved much of its success by using a variety of TRICS (The Representations It Crucially Supposes). They believed that, together, these TRICS led to a cryptoembodiment of rules with the connectionist net. These TRICS all relate to the design and interpretation of the Wickelfeatures in the verb learning model. L&B complained that the selection of 260 particular Wickelfeatures from a possible set of 1210 involved a variety of decisions that tended to reconstitute traditional segmental phonemic information and particular phonological patterns relevant to the learning of the past tense. L&B used this observation to argue that connectionist architectures necessarily contain cryptoembodiments of rules.

11. **The cryptorule problem**: The selective development of an input representation can lead to the cryptoembodiment of rules in PDP nets. See section 5.1 in L&B.

2.3. Additional problems

There are two additional problems with the R&M verb learning model that were not raised by either P&P and L&B.

12. **The early noise problem**: In the R&M model, many epochs of training were required before the model would output phonological forms close to those of recognizable words. There is indeed an early period in child phonology when words have a very indistinct shape. However, when children are working out the various forms of the verb, most of these phonological limitations have been overcome. To be sure, early past tenses include many overregularizations and mismarkings, but these errors are typically modifications of the basic form of the verb, rather than phonologically inarticulate forms.

13. **The direct access problem**: The R&M model works by converting one phonological representation into another phonological representation. This mode of access takes a “basic” form and uses it to find a “derived” form (Bybee, 1985). This is certainly one of the ways in which we can access the correct past tense of a verb. However, intuition, theory (MacWhinney, 1978), and experimentation (Stemberger & MacWhinney, 1986) all suggest that we can also access derived forms directly. In other words, we learn that “ran” means “running in the past” and use this knowledge to access “ran” directly without starting off at “run.” It is not yet clear how connectionist models should be configured to simulate direct access in a theoretically interesting way.

3. **An analysis of the problems**

In this section we will examine each of the 13 problems. We will first examine the way in which the R&M model addressed the issue. Then we will present the criticisms of the R&M treatment and examine the extent to which they are valid. Finally, we will indicate how the problem will be tackled in the new model.

3.1. **The u-shaped learning problem**
Both P&P and R&M characterized morphological learning in children as a three-stage process. This three-stage development was described by P&P (pp. 136-7) in the following terms:

In the first stage, children use a variety of correct past tense forms, both irregular and regular, and do not readily apply the regular past tense morpheme to nonce words presented in experimental situations. In the second stage, they apply the past tense morpheme productively to irregular verbs, yielding overregularizations such as “hitted” and “breaked” for verbs that they may have used exclusively in their correct forms during the earlier stage. Correct and overregularized forms coexist for an extended period of time in this stage, and at some point during that stage, children demonstrate the ability to apply inflections to nonce forms in experimental settings. Gradually, irregular past tense forms that the child continues to hear in the input drive out the overregularized form he or she has created productively, resulting in the adult state where a productive rule coexists with exceptions. (Berko, 1958; Brown, 1973; Cazden, 1968; Ervin, 1964; Kuczaj, 1977; 1981).

There are five major empirical claims involved in the P&P and R&M analysis of a three-staged u-shaped developmental curve.

1. **Early correct usage.** There is an early time when children make correct use of both regular and irregular past tense forms. It is this early period of correct usage that gives rise to the first segment of the u-shaped curve.

2. **Subsequent overregularization.** Following the period of early correct usage, there is a period of overgeneralization of the regular suffix as in “hitted” and “breaked.”

3. **Across-the-board application.** The overapplications of the regular apply to all verbs that the child is using at the time of the beginning of overregularization. It is this across-the-board application that gives rise to the clear middle part of the u-shaped curve.

4. **Coexistence of correct and incorrect forms.** During the period of overregularization, correct and incorrect forms coexist.

5. **Final correct usage.** Eventually, the incorrect forms for the irregular verbs are driven out and only the correct form is used. It is this final period of correct usage that completes the u-shaped pattern of development.

R&M appear to have had a similar understanding of the nature of the u-shaped curve, viewing the curve as applying across-the-board for verbs of all types. R&M attempted to model this pattern by exposing the model to two types of input. In the first epochs of training, the model was given an input set that was very small and rich in irregular forms. At this point, it showed very little evidence of overgeneralizing the regular rule. Presumably, the failure to overgeneralize the regular rule at this point was due not only to the high proportion of irregulars, but also to the small size of the learning set. (A detailed investigation of the effects of such parametric variations on learning is provided by Plunkett and Marchman (1991)). In the next training phase, R&M shifted the nature of the input radically and included a full complement of regular verbs. This shift led to the onset of overgeneralization of the regular rule. One could argue, along with P&P (pp. 140-42) and L&B (p. 237), that this sort of fiddling with the input data is an illegitimate way of deriving the desired phenomenon.

Before we involve ourselves too deeply in this particular debate, we need to consider a fundamental empirical fact. This is that there is no observational or experimental evidence for an across-the-board u-shaped curve. Detailed studies of morphological learning by Derwing (1979; 1979), Derwing and Baker (1979), Kuczaj (1977; 1981) and
MacWhinney (1974) have shown that the onset of the period of overregularization is quite different for different stems. This is true both in observational and experimental work. P&P and R&M both seem to have been unaware of this literature and the extent to which it called into question some of the earlier more global characterizations of the u-shaped curve. More recently, using the data contributed by Brown (1973), Kuczaj (1977), and MacWhinney to the Child Language Data Exchange System (MacWhinney & Snow, 1990), Marcus, Ullman, Pinker, Hollander, Rosen, and Xu (1992) have also concluded that there is little evidence for an across-the-board u-shaped curve. Instead, the three stages of the u-shaped curve come at different times for different verbs, if they can be detected at all. There are some verbs which show a fairly strong u-shaped pattern and others which evidence much weaker u-shaped effects or none at all. Given both a conservative reading of the earlier literature and the recent reconfirmations of that reading, it seems that both R&M's attempt to model a period of across-the-board overregularizations and P&P's critique of that attempt are now largely irrelevant.

The evaporation of the across-the-board model does not mean that the issue of overregularization disappears from our modeling agenda. The other four pieces of the P&P and R&M analysis of the u-shaped curve are based on solid empirical findings. The phenomena of early correct usage, subsequent overregularization, coexistence of correct and overregular forms, and final correct usage do indeed match up with data on language development (Marcus et al., 1992). As we will see later, these four phenomenon still place important constraints on the shape of a full implementation of the acquisition of morphological systems. Of course, there is no complete disappearance of overregularization in adulthood. Rather, adults continue to produce the same types of errors produced by children, albeit with a much lower frequency (Stemberger, 1990).

3.2. The “ated” problem

The R&M model succeeded in producing errors such as “ated” or “ranned” in which the regular suffix is superfluously added to an irregular stem. Because children also produce errors like this, R&M took the model’s ability to produce these forms as support for its empirical adequacy.

P&P questioned R&M's account of “ated”-type errors by arguing that the mechanism which produced them was demonstrably incorrect. In the R&M verb learning model, the form “ated” was produced by activating both a vowel change pattern and the final suffix pattern. P&P argued that these errors are really produced by feeding the past tense formation mechanism with a form that is already in the past tense. The account offered by Pinker and Prince was developed earlier by MacWhinney (1975; 1978; 1982; 1985) and Kuczaj (1977) on the basis of a line of reasoning suggested by Brown (1973). As MacWhinney and Kuczaj pointed out, the fact that children produce errors such as “ating” or “wenting” is good evidence that children occasionally fail to code the irregular past as clearly past. Reduplications such as “jumpeded” show that this confusion is not always confined to irregular verbs. Both MacWhinney and Kuczaj based their analyses on a comparison of experimentally elicited forms and spontaneously produced errors. They both reported that, when children are asked to produce the past tense directly from the present tense “eat,” errors of the “ated” type nearly totally disappear. This further supports the view that at least some of the productions of “ated” are derived from “ate.”

The claim that children sometimes produce “ated” by treating “ate” as input to the past tense formation process can hardly be used to argue against the R&M model, since it would have been an easy matter for R&M to have captured this pattern by simply feeding “ate” as an input to their model. Indeed, our new simulation uses exactly this approach. The real
issue here is not the mechanism for the production of “ated,” but the nature of the semantic developments that must occur for the child to understand the precise nature of the relation between “eat” and “ate.” For an insightful discussion of this problem see Kuczaj (1977, pp. 67-80).

### 3.3. The “hit-hit” problem

The R&M simulation showed fewer overregularizations for irregular verbs that end in /-t/ or /-d/ than for verbs that end in other sounds. In other words, errors such as “cutted” and “hitted” are rarer than errors such as “singed” or “ranned.” P&P (p. 145) attribute the tendency to avoid errors such as “hitted” to “misperception.” According to this analysis, when children and adults perceive the final /t/ on these verbs, they think of it as a past tense marker and then block additional application of the past tense suffix. P&P contrast their misperception account with the R&M account which focuses on the cohesiveness of the “hit-hit” pattern among a subgroup of about 20 irregular verbs. Both P&P and L&B note that certain properties of R&M’s Wickelphonology tend to emphasize the coherence of this subgroup.

In order to understand what is at stake here, we need to examine a variety of contrasting approaches to this problem. The first serious discussion of this problem was presented by Slobin (1971) who claimed that verbs like “hit” provide a “hook” to the regular rule. Slobin apparently thought that children could use this hook to block the repeated application of that rule. A somewhat more symbolic account for this phenomenon is the “affix-checking” account advanced in MacWhinney (1978) and later elaborated by Shattuck-Hufnagel (1979). The “affix-checking” account postulates that, during the production of inflected forms, a filtering mechanism applies to check whether a form appears to end in the required suffix. If it does, further suffixation is suppressed. This mechanism would detect the presence of a final /t/ on “hit” and then block the further addition of -ed.

A major problem with MacWhinney’s affix-checking account was that the process of affix checking does not function in a uniform manner. Sometimes even adults produce “hitted” and “ated.” It is also true that even adults use forms like “pat” and “run” when past tenses are required. In more morphologically complex languages such as Hungarian, the number of affixes that must be checked becomes quite large and one has to postulate a separate checker for each affix. Facts like this led Stemberger (1981) to question the view of affix-checking as a single rule-like process and to think of it instead in terms of interactive activation (McClelland & Rumelhart, 1981) of each activated affix with particular output forms. In Stemberger's model, one type of connection runs between the past tense node and a node for the phonological pattern marking a final dental consonant preceded by a vowel. If a final dental consonant is activated, the activation of the past tense node is decreased. This should lead to avoidance of overmarking and to acceptance of words like “fret” as already marked for past tense. Another set of connections could run between the node representing the past tense affix and various irregular verbs. Stemberger's interactive view was elaborated by Menn & MacWhinney (1984), Bybee (1985), MacWhinney and Anderson (1986), and Stemberger and MacWhinney (1986) using crosslinguistic data, experimentally induced speech errors, and naturalistic speech error data. These authors all assumed a localistic connectionist account in which the partial activation of an output schema was able to then suppress further activation of the past tense morpheme. The basic idea behind all of these analyses is that past tense marking requires that the past tense form “end in a t or d.” Groups such as “brought/caught/fought/wraught/taught/sought” or “bent/lent/sent/rent” illustrate the extent to which having a final /t/ or /d/ is an important output condition for past tense forms.
Given the fact that the misperception account has been offered as a prime example of the role of interactive activation in morphological processing, it is strange to find P&P offering this as their own symbolic account. Perhaps P&P are inclined to accept localist interactive connectionism, but draw the line at the distributed connectionism of R&M’s verb learning model. Alternatively, they may really have intended to embrace the earlier, more symbolic account of this phenomenon of the type proposed by MacWhinney (1978). As Stemberger (1990) notes, it is difficult to figure out exactly where P&P stand on this issue.

In the verb-learning model, the avoidance of “hitted” arises from a match between the form of the stem and the form of the target past tense. Stems with a final /t/ will have one or more Wickelfeatures that will match those used in regular past tense forms. These connections will lead to early production of “hit” as a past tense form, since at least part of the form fits in with the regular pattern, as well as with the majority of verbs in the irregular subclasses. Since the various alternative forms of the past tense are in competition, activation of the final /t/ unit will tend to suppress activation of the syllabic past tense. For verbs like “bend,” the network must also learn to activate the unit for final /t/ and not the units for the syllabic suffix. Of course, this configuration of weights for particular units will fail in the case of verbs like “want,” since they require the full suffix. For these verbs, the network has to learn not to treat the final /t/ as a suffix. The new simulation we will present behaves much like the R&M simulation by allowing acceptance of stem final /t/ as one of the ways of “marking” the past tense.

3.4. The “algalgal” problem

The next four problems raised by P&P and L&B involve the system of Wickelphonology used by R&M. These problems are addressed in the new simulation by using a representation based on autosegmental phonology.

P&P (p. 97) note that the Wickelphonology of the verb-learning model makes it impossible to provide a unique form of representation for words like algalgal “ramrod straight” and algal “straight” in the Australian language Oykangand (Sommer, 1980). For example, the first and second /alg/ triplets would be encoded by a single set of Wickelfeatures, as would the /lga/ and /gal/ triplets. Thus, given an activated set of Wickelfeatures, one would never know if the word was algal or algalgal.

This is a weakness in the system of Wickelphonology used in the R&M model, but there is nothing in connectionism per se that requires that networks use Wickelphonology. The new simulation uses a new type of feature/slot representation that avoids the “algalgal” problem.

3.5. The “slit-silt” problem

P&P (p. 98) also note that the Wickelphonology of R&M makes it impossible to represent the fact that words like “slit” and “silt” are similar in many ways. In the framework of Wickelfeatures, “slit” and “silt” share no features in common. This weakness in representation used in the verb-learning model is corrected in the new representation. In the new representation, the vowel is always in the same slot for both words. The /t/ is in the same slot in the right-justified representation, and the /s/ is in the same slot in the left-justified representation. Otherwise, phonemes end up in different slots but with the same features. Thus, the feature/slot representational system captures both the similarities between these words and also the important differences.
3.6. The “brag-grab” problem

P&P correctly note that the verb-learning model is capable of learning alternation patterns not used in any human language (p. 100):

“A quintessential unlinguistic map is relating a string to its mirror image reversal (this would relate pit to tip, brag to grab, dumb to mud, and so on); although neither physiology nor physics forbids it, no language uses such a pattern. But it is as easy to represent and learn in the RM pattern associator as the identity map. The rule is simply to replace each Wickefeature ABC by the Wickefeature CBA.”

P&P could be claiming either that no single cell of a conjugational or declensional paradigm can use the mirror-image alteration or they could be claiming that no complete paradigm can use it. Presumably they are making the former claim, since any marking that is used uniformly across a paradigm ends up not being a marking at all. The new simulation tests this by examining a language which is just like English, except for the fact that the past tense is formed by the mirror-image transformation. In this language, the rest of the conjugational paradigm uses the standard mapping.

It is not entirely clear that the mirror-image transformation is unlearnable. The fact that languages tend to avoid such transformations may have more to do with the need to avoid morphological leveling and homonymy, rather than any inherent processing problem. Moreover, it appears to be easy to learn to transpose the onset with the coda within a monosyllabic stem. Converting “run” to “nar” or “take” to “kate” is a fairly easy matter. However, full transposition of bisyllabic and trisyllabic stems is intuitively much more difficult. It could be that there is a fundamental learning problem even for monosyllables, but this would require an empirical demonstration.

3.7. The phonological regularities problem

Linguists believe that an important criteria against which any model should be judged is its ability to capture “significant generalizations.” At various points, P&P and L&B argue that the verb-learning model fails to capture certain significant generalizations about English sound structure. P&P note (p. 91) that an English speaker who knows that “Bach” should be pronounced as /bax/ would also automatically realize that the past tense of the neologistic verb “to Bach” would be /baxt/ and not /baxd/ or /baxd/. P&P claim that the R&M model would have trouble producing /baxt/ because it has no clear featural representation of the English sound system. This is another problem that can be addressed merely through a change in the phonological representation. The new model can deal effectively with this problem, because it includes a clear segmental feature representation.

P&P also criticize the verb-learning model for failing to capture certain cross-suffix regularities (pp. 102-108). They note that the same voicing assimilation and vowel insertion patterns that select /t/, /d/, or /l/ as past tense suffixes also operate for the nine homophonous English suffixes that take the form /s/, /z/, or /l/. P&P complain that the R&M model is in principle incapable of capturing these cross-suffix regularities, since it must devote a separate network or module for each suffix. It is important to remember that one of the chief goals of connectionism is to capture the interactions between data types, not to establish a set of encapsulated modules. The new simulation we will present in this paper shows that there is no reason that a connectionist model should require separate modules for each English suffix. The fact that the R&M treated only a single alternation is
a limitation, not of the connectionist enterprise, but of that particular implementation and the machines upon which it was run. The new simulation uses a single network for forming all of the forms of the English verb, including the present (-0), the past (-ed), the participle (-ed and -en), the progressive (-ing), and the third person singular present (-s). For example, the verb "sing" has the forms "sing," "sang," "sung," "singing," and "sings." We will see that this network succeeds in capturing many of the phonological regularities considered important by P&P. In fact, it is precisely because a single module can handle a variety of phonological alternations that it is capable of capturing these linguistically significant generalizations.

3.8. The homophony problem

P&P (section 4.3) note that the verb learning model cannot handle the "elementary" problem of homophony. Consider a pair of verbs such as "ring" and "wring." The past tense of "ring" is "rang," but the past tense of "wring" is "wrung." Moreover, if we use "ring" in the secondary sense of "to form a ring about something," the past tense is "ringed," rather than "rang." Since the verb-learning model takes a single phonological form as its input, it will not know when to produce "rang," "wrung," or "ringed."

P&P maintain that the basic problem here is that PDP nets are incapable of representing lexical identity (p. 174). However, capturing such distinctions in a PDP network is a relatively trivial matter. Taraban, McDonald, and MacWhinney (1989) showed how a net for German morphology could indeed represent the lexical identity of forms simply by coding nouns in terms of a set of semantic features. Doing this with a PDP net is simple and straightforward, as Taraban et al. have shown. The real problem is one of practical computational limitations and the time required to construct a veridical semantic analysis for the lexicon of a language. When a large number of words are involved, a great deal of semantic features must be used to disambiguate them all. Adding a large set of semantic features to characterize the meanings of individual words increases the size of the network and markedly slows down simulation runs. When this is done, a vast amount of network resources must be devoted to the fairly arbitrary mapping from semantics to phonetics. Forcing the network to devote resource to acquiring this mapping drastically slows simulation runs in that network resources must now be dedicated to both phonological and lexical mappings.

Our solution to this problem in the new simulation is to examine the homophony problem within the context of a small lexical set. We do this in a secondary simulation which supplements the phonological architecture of the main simulation with additional nodes for semantic features.

3.9. The convergence problem

P&P (p. 125) and L&B (P. 218) note that the verb-learning model makes errors in computing the past tense forms of a large percentage of the words it is tested on. Indeed, there are some forms for which it produces no past tense at all, despite the use of a binding network to bring productions into accord with real lexical items. When tested with 72 regular verbs, the network produced incorrect past tense forms for 20 of these verbs. Of these 20 errors, there were six with no past tense response at all. P&P and L&B are correct in criticizing this level of performance, particularly given the reliance on the binding network to improve performance. They correctly note that a major source of the problem is the fact that the network does not have a way of making modifications to a present tense form to produce a past. However, this problem is specific to the implementation chosen by
R&M and not basic to the connectionist conceptualization. The new simulation treats the learning of the past tense as the development of a set of connections for modifying the present tense form. This assumption, along with several other implementational changes, allows the new simulation to reach a much higher level of performance.

3.10. The regular pattern problem

P&P fault R&M for using a single mechanism to process both regular and irregular verbs. P&P want to make a fundamental distinction between “fuzzy families of memorized exceptions” and “formal rules.” The irregular patterns survive only through relying on family resemblance, high frequency, and gang effects, the general pattern survives through its formal accuracy and general applicability.

This contrast between the regular pattern as a formal rule and the minor patterns as fuzzy families of memorized exceptions involves a number of questionable assumptions. Perhaps the most serious problem with this view is the claim that all phonological markings have some preeminent regular pattern. Consider a system such as the marking of plurality on the German noun. German plurals can be formed using -en, -er, -s, -e or zero as endings, along with possible vowel ablauting. None of these five possible suffixes is statistically predominant (Köpcke, 1988); none can be characterized as being “the regular ending.” In a situation such as this, there is simply no regular pattern at all. Are we to draw some sharp line between English and German speakers by claiming that only the former evidence “rule-governed” behavior?

Somewhere between the German plural and the English past tense is the Hungarian plural. There we find a set of five competing forms -k, -ők, -ék, -ők, and -ak. Many highly “irregular” verbs take the -ak suffix and also undergo changes in the vowels and consonants of the stem (MacWhinney, 1978). However, not all irregulars take the -ak suffix. If the stem ends in a consonant, the -ok, -ek, -ők triplet can be viewed as the “regular” pattern. However, the selection of one of these forms is based on a set of vowel harmony patterns which have now been shown to be full of fuzziness and leakage (Ringen & Kontra, 1989) of precisely the type that P&P want to ascribe to the irregular classes in English.

The words “regular” and “irregular” are useful ways of referring to the two ends of a continuum. However, we should not be misled by their usefulness into identifying “regularity” with rules and “irregularity” with memorized forms. Even in English, the regular class leaks on its borders. One of the salient properties of the regular class is the fact that both the past tense and the past participle are formed by suffixation of the -ed form. However, the regularity of this paradigmatic neutralization is incomplete. Consider the verb “prove.” Is this verb regular or irregular? In the past tense, it takes the regular form “proved.” From this we would judge it to be a regular verb. However, the past participle takes the form “proven” which has a decidedly irregular shape. If we simply assume that “prove” is an irregular, we then have to explain why it is that it acts like a regular in the past tense.

P&P might argue that these cavils miss the fundamental issue. What is crucial is the way in which the regular pattern imposes its sway on recalcitrant verbs. The regular pattern operates less through family resemblance than do the minor patterns. The crucial example (p. 117) is the form “kneelt” which first appeared as a substitution for “knealed” in the 19th century, presumably under the influence of irregulars such as “dealt,” “spelt” and “felt.” P&P argue that the fact that “heal,” “peal,” and “seal” were not able to keep “kneal” away from the clutches of the irregular verbs indicates that the regular pattern does not work
through family resemblance. Rules do not need to pay attention to the phonetic substance of the items to which they apply, because they apply by default to all items. Marcus et al. (1992) have provided further support for this analysis by showing that, during language learning, irregular verbs are protected from the regular pattern by neighborhoods of similar sounding irregulars, but are not attracted to regularization by neighborhoods of similar sounding regulars. In other words, regular items appear incapable of launching a “gang effect.”

P&P seem to have the facts correct here. The question is how to interpret these facts in terms of the debate between connectionism and symbolism. Is there anything in the original R&M analysis that commits connectionism to viewing the regular pattern in terms of family resemblance or gang effects? The answer is that R&M are not arguing against rules, but for networks.

“We suggest instead that implicit knowledge of language may be stored in connections among simple processing units organized into networks. While the behavior of such networks may be describable (at least approximately) as conforming to some system of rules, we suggest that an account of the fine structure of the phenomena of language use can best be formulated in models that make reference to the characteristics of the underlying networks.” (Rumelhart & McClelland, 1987, p. 196)

What P&P seem to fail to understand is that, within connectionist networks, general patterns are less dependent on gang effects than are limited patterns. Limited patterns can only exist if they are supported by very tightly defined neighborhoods or gangs. Both the shape of the output past tense phonology (Bybee, 1985) and similarities in the beginning of the stem determine the neighborhoods of minor patterns. The English regular past tense, on the other hand, defines its neighborhood entirely on the basis of the final segment of the stem. Because it only pays attention to the position and voicing of that final segment, it is not prepared to defend against incursions from marauding minor patterns. However, when those minor rules have lost some of their own internal coherence, one of their members may simply fall out of the gang into the net of the waiting general pattern.

Given the fact that both the connectionist and the symbolist accounts seem to be happy with the contrast between minor patterns and major patterns, it will be important to see how a full symbolic simulation deals with this issue and how it succeeds in accounting for learning in languages where the major patterns show more leakage than in English.

3.11. The cryptorule problem

L&B claim that, insofar as the verb learning model works at all, “it actually confirms the existence of rules as the basis of natural language.” They attempt to support this claim by showing that the way in which R&M selected Wickelfeatures involved a variety of decisions that tended to reconstitute traditional segmental phonemic information. The new simulation takes a very different approach to phonological representation. Instead of using Wickelfeatures, it uses traditional phonological features within a simple feature/slot framework.

Given the fact that L&B claim that acceptance of a featural notation or even crypto-acceptance of a featural notation implies acceptance of rules, one might think that the L&B criticism should apply equally well to the new simulation. However, the new simulation shows that the L&B criticism was focused not on the overall connectionist conceptualization, but on the particular implementation presented by R&M. L&B need to
develop an argument that links a reliance on phonological features to an acceptance of a rule-based phonology. However, they never develop this argument. This surprising gap in their argumentation was probably caused by their attempt to show that R&M were indeed relying on something like standard phonological features. The new simulation relies on no “TRICS” other than those accepted in standard phonological analysis. To extend their TRICS argument to the new simulation we will present here, L&B would have to show that acceptance of any featural coding involves acceptance of a production-system architecture. This would be a hard position to defend. Indeed, the connectionist architecture itself shows that there is an alternative way of going from a featural representation to the learning of patterns of word formation. In a sense, that is the whole point.

3.12. The early noise problem

There are two additional problems with the R&M simulation that were not discussed by either P&P or L&B. The first of these is the early noise problem. As we noted earlier, the initial past tense verbs produced by the R&M model often have a shape very unlike that produced by two-year-old children. Instead, they look like the words produced during the transition from babbling to speech in the second year. If the R&M simulation were designed to model early phonological development, this would not be a problem. However, the verb-learning model is not a simulation of early development, but of the learning of the past tense as a modification of the phonological form of the present tense. Because of this, it is safe to say that the R&M simulation suffers from a period of “early noise.”

The model that will be presented here addresses this problem by taking the “basic” form of the present as a starting point for the formation of the “derived” forms of the past, the participles, and the third singular. In the new model, this is a relatively easy thing to do. Basing the derived forms on the basic helps to speed learning, capture developmental facts, and promote convergence.

3.13. The direct access problem

A fundamental limitation of the verb-learning model was its inability to represent the direct access to a past tense form from a meaning. If we can go from the meaning underlying “dog” to the phonological form /dOg/, there is no reason to believe that we cannot go from the meaning underlying “took” to the phonological form /tVk/. However, the verb-learning model does not allow this type of direct access.

The new simulation does not solve the direct access problem. It would be possible to represent direct access by simply setting up a series of “localist” nodes that would map a particular meaning to a particular phonological pattern and vice versa. A system of this type is easy to construct. However, it would fail to capture certain key aspects of phonological and semantic development. Alternatively, words could be accessed through lexical subnets which focus on selecting between words with closely competing meanings. This second possibility is discussed in MacWhinney (1990). The construction of even a toy network of this second type is a fairly major task and clearly outside of our current concerns. For the moment, we will have to simply accept the fact that resolution of the direct access problem is a major task for future connectionist models of morphological processing.
4. The new simulation

The new simulation involves a new input representation, a new output representation, a new network architecture, a new learning algorithm, and a new input corpus.

4.1. The input representation.

The new simulation uses 214 phonological input units. Each input phonological unit codes for two types of information. One type of information is the position of the unit in the syllabic template. The other is the value of some particular distinctive feature. In order to understand this coding system, we need to examine both the syllabic template and the feature coding system.

The syllabic template is designed to reflect current approaches to autosegmental phonology. There are two parts of the template: a left-justified part and a right-justified part. The left-justified template takes the form CCCVVCCCVVCCCVVCCC, where C stands for consonant and V stands for vowel. This pattern codes a full trisyllabic structure in left-to-right fashion. The right-justified template takes the form VVCCC. This pattern only represents the coda of the final syllable. Vowel nuclei are composed of up to two segments and consonantal clusters are composed of up to three segments. If a particular segment is not actually present in a word, its features are simply left off.

The featural coding system is based on a fairly conventional distinctive feature representation for English. No Wickelphones or Wickelfeatures are used. The 14 vowels of English are represented by eight distinctive features. Since there are six possible vowel slots in the trisyllabic pattern and two in the coda pattern, there are a total of 64 units dedicated to vowels. The 22 consonants of English are represented by ten distinctive features. Since there are twelve possible consonantal slots in the trisyllabic pattern and three in the coda pattern, there are a total of 150 units dedicated to consonants. Together, the 150 consonantal units and the 64 vocalic units yield a combination of 214 feature/slot units. The feature breakdown for the 14 vowels and 22 consonants in English is as follows. We used the UNIBET (MacWhinney, 1991) to assign single-letter ASCII codes for each phoneme.

The vowel features were: front, center, back, round, high, middle, low, and diphthong. Using these eight features, the eleven monophthongs and three diphthongs of English were coded in the following way.

\[
\begin{align*}
\text{i} & \text{ front} \\
\text{I} & \text{ center, high} \\
\text{e} & \text{ front, middle} \\
\text{E} & \text{ front, middle, low} \\
\& & \text{ front, low} \\
\text{u} & \text{ back, round, high} \\
\text{U} & \text{ center, back, round, high} \\
\text{o} & \text{ back, round, middle} \\
\text{O} & \text{ back, round, low} \\
\text{a} & \text{ center, low} \\
\text{6} & \text{ center, middle} \\
\text{1} & \text{ front, back, round, high, low, diphthong} \\
\text{2} & \text{ front, back, round, high, middle, diphthong} \\
\text{3} & \text{ front, high, low, diphthong}
\end{align*}
\]
The consonantal features were voiced, labial, dental, palatal, velar, nasal, liquid, trill, fricative, and interdental. The 22 consonants of English were coded in the following way:

- p labial
- t dental
- k velar
- b voiced, labial
- d voiced, dental
- g voiced, velar
- m voiced, labial, nasal
- n voiced, dental, nasal
- N voiced, palatal, velar, nasal
- l voiced, dental, palatal, liquid
- r voiced, dental, trill
- f labial, dental, fricative
- v voiced, labial, dental, fricative
- s dental, fricative
- z voiced, dental, fricative
- S palatal, fricative
- Z voiced, palatal, fricative
- j voiced, palatal, liquid, fricative
- h velar, fricative
- w labial, liquid, fricative
- T dental, fricative, interdental
- D voiced, dental, fricative, interdental

In the left-justified representation, segments are packed to the beginning of the CCCVVCCCCVVCCCVVVCC pattern. For example, the word /bEt/ is fit into the left-justified template in this way:

bCCEVtCCVVCCVCCC.

The same word was fit into the right-justified coda template in this way:

VECCt.

The /t/ of /bEt/ stands out as “final t” in the right-justified template, whereas in the left-justified template it is “the initial consonant in the first post-vocalic cluster.” In the terminology of L&B, the use of these two views upon the syllabic structure of the word can be thought of as TRICS. However, these TRICS are in no way modeling “tricks.”

The basic notion of a set of slot/feature units derives in a motivated manner from the theory of autosegmental phonology (Levelt, 1989; Nespor & Vogel, 1986; Selkirk, 1984). The use of both a right-justified and a left-justified form derives from empirical work on language processing and acquisition that indicates that both children and adults pay attention to the beginnings and ends of words (MacWhinney, 1978; 1985; Slobin, 1973).

In addition to the 214 feature/slot units, the input includes five units dedicated to the five cells in the paradigm for English verbs. There is one unit each for the present, the past, the past participle, the present participle, and the third person singular present. These units can be thought of as simple ways of notating more complex semantic intentions. Thus, if the phonological pattern for “sing” is turned on in the 214 feature/slot input units and the input unit for the past participle is also turned on, the output should be “sung.” However, if the unit for the present participle is turned on instead, the output should be “singing.”

4.2. The output representation.
The output of the network is simply the 168 feature/slot units of the left-justified phonological form. The right-justified coda representation is not included in the output.

4.3. The architecture of the network.

Between the input and output units there are two layers of 200 “hidden” units. These units are called “hidden” because they have no direct interpretation in either input or output terms. Between adjacent layers, every unit of one layer is connected to every unit of the other layer. The model uses two layers of hidden units, because a model which had only one layer of units did not do as well at learning the training set.

The final feature of the network is a set of identity-mapping connections between the left-justified input and the output. The network was designed to treat the learning of the “derived” forms of the verb as modifications of the phonological form of the “basic” present tense. The idea is that the child assumes that the past tense is somehow a modification of the present. This is done by including a set of connections that “copy” the left-justified phonological form of the input directly onto the output. This copying only sets a weak bias on the activation of the output units. This bias can be overcome with learning. Indeed, as we will see, the initial bias is usually overcome within a few trials.

4.4. The learning procedure.

A major difference between the current model and the R&M model is that the new simulation uses the back-propagation learning algorithm (Rumelhart, Hinton & Williams, 1986). This algorithm makes use of layers of hidden units to capture nonlinearities in
problem spaces. The improved learning of the training set found in the new simulation is at least partly attributable to this more powerful learning algorithm.

Training consisted of the presentation of approximately 1.3 million forms. Forms were presented with the actual relative frequencies found in Francis and Kucera (1982), with the most frequent form being produced exactly once per epoch, and the rarest forms about once every 700 epochs. This resulted in the presentation of about 1.3 million forms during the course of training. This is a large number of forms, but it is probably two or three orders of magnitude less than the number of English verb forms heard by a child.

4.5. The input corpus.

The input corpus was derived from the Francis and Kucera (1982) corpus of English word frequencies. The 6949 most frequent verb forms - including present, past, past participle, present participle and third person singular - were first selected as a base set for the simulation. These were derived from 2161 different verbs. Homophones and multiple forms (e.g. past tense for “spit” can be either “spit” or “spat”) were eliminated by extracting the less common forms. All forms which had more than three syllables, more than three consonantal phonemes in a row, or more than two vocalic phonemes in a row were also removed. The remaining 6090 forms, derived from 2062 verbs, represented the corpus of forms used in the simulation. Of these the least frequent 10% of the regulars and the least frequent 10% of the irregulars were extracted and saved for testing the generalization abilities of the network. The remaining 5481 forms were used for training. The training set included these 118 irregular past tense forms: arose, ate, awoke, beat, became, bent, bled, blew, bore, bought, bound, broke, brought, built, burst, came, cast, caught, chose, clung, cost, crept, cuî, dealt, drank, drew, drove, dug, fed, fell, felt, fled, flew, flung, forgave, forgot, fought, found, froze, gave, got, grew, ground, heard, held, hid, hit, kept, knew, laid, led, left, lent, let, lost, made, meant, met, overcame, put, quit, ran, rang, read, rode, rose, said, sang, sank, sat, saw, sent, set, shed, shone, shook, shot, shut, slept, slid, sold, sought, spat, spent, split, spoke, sprang, spread, spun, stole, stood, strode, strove, struck, stuck, swam, swept, swore, swung, taught, thought, thrust, told, took, tore, understood, underwent, upheld, upset, went, wept, wet, withdrew, woke, wore, wound, wove, wrote. This is not the complete set of irregular verbs in English, but it includes the most frequent irregulars.

5. Results

In this section we will examine the ways in which the results of the new simulation address the 13 problems. With simulations of this size it is difficult to save enough data to fully represent the events that take place during training. In our case there were a total of 6090 forms that could have been monitored throughout the 24000 epochs of training. The output for each of these was represented by an array of 168 floating point numbers. Thus about 90 gigabytes would have been required to save the output for every form at every epoch. To make the analysis more tractable, we saved only the output generated for each of the 236 irregular forms in the training set at every 10 epochs. In addition, every 4000 epochs the output generated for all of the 6090 forms was saved to monitor the overall performance of the network as it learned. The basic form in which we examined the data was as strings of UNIBET characters representing the phonemes that appeared in each slot of the output pattern. The current analysis focuses only on the results for the past tense. Within the past tense, the primary focus is on the 118 irregular verbs in the training set. Even for this subset of the results, there are 118 x 2400 forms to examine. To further collapse the data set, we pulled out blocks in which particular responses predominated and then simply
noted trials in which secondary responses occurred against a background of primary responses. Our reports below are based on this final distillation of the database.

The results for the progressive and the third singular present are quite simple to summarize. Not surprisingly, the progressive was fully correct for all forms by the first observation point at epoch 4000. The third singular present was also learned early on for most forms, but there were errors for sibilant final verbs such as “pledge” or “buzz.” The results for the past participle are not easily summarized, since there are many irregular forms and many persistent errors. Indeed, at the end of the simulation, there were a few more errors being produced on the past participle than on the past tense. This finding corresponds with observations such as those reported by Zwicky (1970).

The most relevant results are those for the past tense. First, let us consider some general aspects of the irregular past tense results.

1. Typically, the present tense was produced as output during the first 40 to 80 epochs. This was in a large part due to the identity-mapping biases set on the output units from the beginning of training.
2. For zero-marking irregulars such as “cut” or “hit,” the period of direct reproduction extended as far as epoch 200 before the first attempts were made to produce overregularizations such as “cutted” or “hitted.”
3. For forms with vowel changes, the network spent a period from about epoch 200 to about epoch 3000 exploring alternative vowel shifts and modifications.
4. Concurrently with explorations of vowel shifts, the network explored use of final /t/, /d/, and /Id/.
5. For most of the forms, correct performance was stabilizing by epoch 3000. However, a few late vowel errors and overuses of final /t/, /d/, or /Id/ can be found up to around epoch 6000.
6. For most verbs, the last 10,000 epochs are error free.
7. There were three forms that were only learned in the last 10,000 epochs. These were “bled” at 14780, “brought” at 15810 and “thought” at 15820.
8. Only eleven low frequency irregular forms remained unlearned at the end of training.
9. In the testing done at the end of the simulation, nine of the thirteen untrained past tense irregulars were missed. Six of these were simply mistaken as other classes of irregulars or as regulars, as the network had no way to know for sure to which class they belonged.

Performance on the regular past tense was perfected about midway through the simulation. By the first test point it was already over 99% correct. In overall terms, the simulation obviously did very well at its task of learning the past tense. These results are encouraging, but how well was the model able to address the 13 specific problems we examined earlier?

5.1. The u-shaped learning problem

We noted earlier that u-shaped learning for verbs involves four observable components -- early correct usage, subsequent overregularization, coexistence of correct and incorrect forms, and final correct usage. The new simulation correctly models the last three components.

1. Overregularization. Usually, overregularization begins shortly after the first 50 or 100 epochs. For nearly every one of the irregular verbs, except for zero-marked
pasts such as “hit” or “cut,” there is a clear early period of overregularization. This is not an across-the-board phenomenon, since some verbs show much more overregularization than others and some show it for a much longer period than others.

2. **Coexistence of correct and incorrect forms.** There is then a period of between 300 and 3000 epochs for almost all irregular verbs where the correct form coexists with overregularized forms. For a verb like “bent” this period ends by epoch 2800; for a verb like “arise” it extends to epoch 8010.

3. **Final correct usage.** For nearly all of the verbs, there then follows a long period of final correct usage. It is important to note that, once the network reaches correct performance, deviations from the correct model are fairly infrequent. There are indeed late occurring errors against a background of correct performance, such as the uses of “binded” for “bound” on epochs 1950 and 2300. However, when we realize that the learning went on for 24000 epochs, it is clear that much of this period was spent in error-free performance for this verb. In this sense, the model behaves fairly deterministically, reflecting the deterministic nature of the learning algorithm. If we were interested in simulating adult speech errors, we would have had to employ a non-deterministic model.

However, the simulation fails to capture the fourth component of the u-shaped pattern -- early correct usage. Verbs like “hit” are correct from the beginning, but this is simply because they are zero-marked. Regular forms also show early correct usage. Within the irregulars, there are a few very early uses of forms such as “ate” or “ran,” but these early correct uses do not occur in the very first epochs, but rather in the period of the first 500 epochs when the network is exploring a variety of possible past tense forms. At the very onset of learning, most of the past tense output for irregulars is either the present tense or a regularization, rather than the correct past tense. However, real children often produce correct irregular forms early on. In this sense, the model fails to capture this period of correct early production of irregular forms.

A decision was made not to attempt to model the stage of early correct production of irregular forms. This early learning of the past tense presumably works not by mapping from the sound of a basic form to the sound of a derived form, but by direct access from meaning to sound. As we will note in our discussion of the direct access problem, this type of early learning could be modelled in a “brute force” way by simply setting up a set of rote-memorized lexical units. However, doing this would not illuminate the process. It seems better, at this point in the game, to be honest about the fact that there is not yet a good way of dealing with direct access within this model.

### 5.2. The “ated” problem

P&P argue that forms such as “ated” arise when the child confuses “ate” with the present. Capturing this type of processing in this network was extremely easy. We simply constructed a second set for testing generalization that was composed of four irregular past forms. The verbs were “ate,” “broke,” “ran,” and “bought.” The results are not particularly surprising. The network produced as output the forms “ated,” “broked,” “ranned,” and “boughted.” This was the case for “ate,” “ran,” and “bought” at the first test point (epoch 4000). “Broke,” however, yielded “broke” as its past tense in this particular type of generalization testing until the test at 12000 epochs, after which point it produced “broked.”

At the same time, the network produced forms such as “ated” and “ranned” without using the past tense as input. As in the R&M model, these forms are produced by simultaneous
application of vowel change and suffixing patterns to a single input. For example, the input “eat” undergoes both a change of the vowel and an addition of a suffix to produce “ated.” Overmarked forms of this type are quite numerous in absolute terms, although they constitute only a small percentage of the total output. What these two observations show is that overmarking can easily be produced by connectionist nets either through simultaneous application of marking processes or through inflection of a base form that is already marked. This already marked base form has exactly the same status as the other base forms that serve as input to the model.

5.3. The “hit-hit” problem

Like the R&M model, the new simulation produced fewer overregularizations for zero-marking verbs like “cut” and “hit” than for other irregulars. Indeed the zero-marking class was learned quite quickly with very few errors. For example, “burst” had only three “bursted” errors, and only seven errors in total. The verb “cast” had only one “casted” error after epoch 1950; “cost” had none after epoch 1340, and “cost” had none after epoch 210. Perhaps the most dramatic evidence for the power of what Menn and MacWhinney (1984) called the “repeated morpheme constraint” is the fact that the verb “cut” had only three errors in 24000 epochs -- two of these were uses of the overregularized “cutted” on trials 170 and 210. When we compare this virtually error-free learning of “cut” with the significantly greater difficulties the network had in learning an alternation such as “deal - dealt” which combines a vowel change with addition of final /t/ after a liquid, it is clear that verbs of the “hit” and “cut” type are much easier to learn than other irregulars. This finding matches well with the developmental data discussed earlier, as well as the theoretical and empirical work of MacWhinney (1978), Menn and MacWhinney (1984), Stemberger (1981), and Stemberger and MacWhinney (1986). In the present model, the fact that “hit” and “cut” end in a /t/ is given particular weight by the identity mapping connections which set up “hit” as the network's first guess regarding the identity of the past tense of “hit.”

5.4. The “algagal” problem

As we noted earlier, the “algagal” problem disappears in the new simulation. The phonological representation used in the new simulation guarantees unique representations for different sounding words. The left-justified representation of “algagal” in the Oykangand language is CCCaVlglVlCCVVCC, whereas that of “algagal” is CCCaVlglVlglVaVlCC. Since these different words have different representations in the new model, it is fair to say that the “algagal” problem is solved by the new simulation.

5.5. The “slit-silt” problem

The “slit-silt” problem also disappears with the new simulation, because phonologically similar words now have more similar representations. In the case of “slit” and “silt” the /i/ appears in exactly the same slots for both words. The final /t/ also appears in the same slot in the right-justified coda representation. The matches are indicated below with the letter “x.” Thus, the feature/slot representational system captures some of the similarities between these words as well as some of the differences.

```
“slit” left-justified: s1CiVtCCVCCVCCVCC
“silt” left-justified: sCCiVtCCVCCVCCVCC
x
```
As an implementational matter, it is not entirely clear whether the particular featural representation selected was fully adequate. For example, an occasional error for the verb “catch” was /k0tSt/ or “caughtshd”. If the affricate /tS/ had been coded as a single phoneme, this rather improbable error might never have been produced. It is also possible that the slow learning of some of the vowel change patterns could have been speeded by the selection of a richer set of features for the vowels. However, resolution of these details must await further work.

5.6. The “brag-grab” problem

We ran an auxiliary simulation that was identical to the basic simulation, except that in the auxiliary simulation the past tense was formed by reversing the order of segments in the present tense. For example, the past tense of “brag” was “grab” and the past tense of “trickle” was “lkirt.” All past tenses were formed according to this absolutely uniform mirror-image construction scheme. This transformation greatly impaired the performance of the network on the past tense. After 24000 epochs of training, only 15% of the forms were correct. Apparently, the network cannot learn an alternation of this type.

5.7. The phonological regularities problem

The new simulation addresses the phonological regularities problem in a variety of ways. First, the change to a feature/slot representation improves its ability to model generalization across the phonological inventory. To demonstrate this, the simulation was also given the verb “bach” /bax/ as a generalization test and correctly produced the past tense form “bached” /baxt/. Second, by including multiple forms of each verb in training, the simulation was able to demonstrate cross-paradigm regularities. For example, the progressive of “accompanying” was occasionally formed as “accompanyng.” The deletion of the initial vowel of the progressive “-ing” suffix appears to be a result of paradigmatic pressure from similar deletion patterns occurring in the other suffixes. Eventually, the network learns to counteract this pressure, but the interesting fact is that the new simulation shows the presence of such phonological regularities. Other evidence of the impact of phonological regularities is provided by interpenetrations of past participle forms and irregular pasts, as in “sunged” for “sang” or “aten” for “ate.” For an interesting report on a persistent case of such interpenetration in a single child, see Zwicky (1970).

5.8. The homophony problem

A small auxiliary simulation was run to demonstrate the ability of a network to acquire past tense forms for homophones. The verbs used were: ring, wring, jump, want, run, talk, pull, answer, and paste. The network architecture was supplemented by the following 21 semantic features: action, auditory-result, cause-contact, circle, completive, high-pitch, internal-state, object-gap, object-state, object-thing, positional-change, response, sharp-onset, speech-act, surround, torque, use-of-hands, use-of-feet, vertical direction, volitional, and whole-body-motion. The exact feature coding used was as follows.

ring-1    action, auditory-result, high-pitch, object-thing, sharp-onset
ring-2    action, circle, completive, object-thing, positional-change, surround
The results of this simulation were quite simple. The network learned to produce “rang” as the past tense of “ring-1,” “ringed” as the past tense of “ring-2,” and “wring” as the past tense of “wring” within 2400 epochs. These results indicate that nets of this type can readily resolve homophony.

5.9. The convergence problem

By epoch 4000, all but seven of the regular past tense forms were being produced correctly. By epoch 8000, all but one of the regular past tense forms were correct. By epoch 16,000, all of the regular past tense forms were correct. By the end of training at epoch 24,000, errors were being made on only eleven of the irregular pasts. Perfect performance on the present progressive was achieved by the first check at 4000 epochs. The third person singular present was perfected by the end of the simulation. Learning of the past participle followed a pattern similar to that for the past. It is clear that the network succeeded at its assigned task of learning the English verb paradigm. If we had allowed the network to run for several additional days and given it additional hidden unit resources, we probably could have reached complete convergence.

5.10. The regular pattern problem

Regular pasts were learned fairly early and without significant error. By the first check at epoch 4000, only seven of the 1059 regular pasts were being missed. This dropped to one miss by the following check at epoch 8000, and by epoch 16000 performance was perfect. This type of learning is exactly what we see in children. When we tested thirteen untrained irregulars at the end of training, four were treated as regulars. Unfortunately, we did not test a similar set of thirteen regulars, but presumably they would have been treated more uniformly as regulars, in accord with the analysis of P&P. It is not clear that any of these results should be viewed as inconsistent with either the connectionist approach or the symbolic alternative.

5.11. The cryptorule problem

L&B criticized R&M for biases in the selection and construction of Wickelphones. Those biases are not present in the new coding system. There is nothing in the featural representation of the current simulation that biases it toward the acquisition of some particular rule. Instead, the representation was chosen to be powerful enough to facilitate learning of different types of rules in different languages. It is true that the representation is capable of expressing something like “final t.” However, there is nothing wrong with that ability and nothing that can be construed as involving either TRICS or “tricks.”
5.12. The early noise problem

The new simulation no longer spends its initial epochs groping toward the ability to produce some recognizable form of the verb. From the first trials, it is producing either the present tense or some variant of the present tense. In this regard the model provides a more accurate characterization of the way in which the two-year-old child works on this task.

5.13. The direct access problem

The new simulation does not solve the direct access problem.

5.15. Summary of results

Let us review in as brief a form as possible how the 13 problems were addressed in the new simulation.

1. **u-shaped learning:** No attempt was made to switch the shape of the input. The network showed the final stages of u-shaped learning, but not correct initial learning. This failure can be attributed to the model's inability to capture the kind of simple rote lexical access that occurs at the earliest stages of word learning.

2. **ated:** The model produced “ated” both through simultaneous application of vowel change and suffix addition and through redundant suffixation of “ate” treated as an input.

3. **hit-hit:** The model showed the correct avoidance of overmarking for final /h/ irregulars. It does this in a natural way, rather than by mere stipulation, as in a more symbolic account.

4. **algagal:** The adoption of a new input representation solved this problem.

5. **slit-silt:** The adoption of a new input representation solved this problem.

6. **brag-grab:** The adoption of a new input representation solved this problem.

7. **phonological regularities:** The modeling of the full verbal paradigm and the adoption of a new input representation solved this problem.

8. **homophony problem:** An auxiliary simulation demonstrated that this is not a problem.

9. **convergence:** The adoption of a new architecture, a new representations, and a new learning algorithm solved this problem.

10. **regular pattern:** It is not clear that this is a problem or even what the exact nature of the problem might be.

11. **cryptorule:** The adoption of a new input representation solved this problem.

12. **early noise:** The use of identity mapping connections solved this problem.

13. **direct access:** This problem has not been solved.
All in all, the new model did quite well. The remaining problems are the model’s inability to handle the direct access problem and the resultant problem with the correct learning during the first stage of u-shaped learning. Given the fact that a mere reimplementation of the original R&M model succeeded in solving so many problems that had been characterized as conceptually “profound,” we are forced to conclude that the critiques of P&P and L&B suffer from a confusion of implementational issues with conceptual issues.

There is reason to believe that the difficulty that the current simulation encountered with producing early correct irregulars such as “went” will be easy to overcome in a somewhat different version of the current architecture. In work that is currently in progress, we have supplemented the current model with a full set of “semantic” encodings for each lexical item. These encodings are simply a set of arbitrarily selected features which serve the function of providing distinctiveness between lexical items. In this new simulation, a new frequency-based selection algorithm is used that guarantees that the earliest epochs have a greater concentration of high frequency forms and later epochs have a greater concentration of low frequency forms. This reflects our belief that children tend to filter out lower frequency forms during the beginning of language acquisition. Together, these changes in the frequency selection algorithm and the addition of “semantic” features allows this new simulation to deal more adequately with the problem of early correct irregulars. It successfully manages to produce “went” and then later goes through a period when “went” competes with “goed.”

6. Is there a better Symbolic Model?

If there were some other approach that provided an even more accurate characterization of the learning process, we might still be forced to reject the connectionist approach, despite its successes. The proper way of debating conceptualizations is by contrasting competitive implementations. To do this in the present case, we would need a symbolic implementation that could be contrasted with the current connectionist implementation. P&P sketch out a piece of such an alternative. It turns out that the model they attribute to Pinker (1984) is in fact the model proposed by MacWhinney (1978) for the acquisition of simple phonological alternations.

According to the model developed in MacWhinney (1978), during the first phase of acquisition, analysis is used to factor out suffixes. A base form is compared with a derived form and any material on the left or right edge is analysed out a new morpheme. For example, by comparing “jump” with “jumped” the past tense suffix is analyzed out. During the second phase of learning, a process which MacWhinney (1978) calls “production acquisition” is used to factor out stem changes. For example, by comparing “sing” with “sang,” a rule or production is formulated with changes /ai/ to /ae/ in the very specific phonological context of /s_N/. Next, if two words undergo the same change, such as “ring” and “sing,” their phonological contexts are superimposed to yield a more abstract context for the production. Each new superimposition increases the strength of the rule. Following the “selectric principle” of Cearley (1974), productions match the input in parallel and compete on the basis of their strengths. MacWhinney develops the explanation in terms of Hungarian stems and suffixes; P&P present an illustration based on the English past tense rule. However, the processes used in the two cases are identical.

The model that P&P are proposing as the main symbolist alternative to the R&M model is the model of MacWhinney (1978). We do not wish to argue that the MacWhinney model was ill-constructed or descriptively inadequate. However, we do not believe that P&P can demonstrate that it is superior to the connectionist alternative we have presented here.
There are certain detailed problems with the earlier MacWhinney account. It did a poor job of capturing the distinction between the regular and the irregular patterns which P&P judge to be so important. In that model, it was often difficult to know whether a new phonological alternation should be treated as a case of an old production or the basis for a new production. If the merger procedure is extremely conservative, certain phonological regularities will not be captured. For example, in Hungarian, this procedure would never merge across vowel harmony types. If the merger procedure is not sufficiently conservative, there can be problems with retreating from overgeneralizations. When it comes to actually implementing the symbolic account, a myriad of detailed decisions must be made regarding the shape of possible input forms, the ways in which rule strengths should be incremented, the algorithm for production matching and conflict resolution, and so on.

Although P&P are willing to accept the further consequences of having adopted the approach of MacWhinney (1978), it is clear that L&B are not. A phonology with strict rule ordering of the type espoused by L&B (p. 203) is clearly incompatible with a parallel rule-matching algorithm of the type proposed by MacWhinney (1978) and elaborated in MacWhinney and Anderson (1986). This comparison between the P&P and L&B position on rule systems is instructive, since it underscores the fact that the symbolist approach yields no single monolithic alternative to connectionism.

7. Summary

We have examined a series of 13 problems raised by the verb learning model of Rumelhart and McClelland (1986, 1987). These problems were presented by P&P and L&B as fundamental flaws in the conceptualizations underlying connectionism. They believed that, by calling attention to these flaws, they could well call into question any application of connectionist models to language processing or learning. To address these problems, we constructed a new simulation using a new input representation based on feature/slot units and two views on input words, a new architecture using identity mappings and hidden units, a new learning algorithm, and an input corpus that included all five cells of the English verb paradigm. Together, these changes led to a vast improvement in the performance of the model. All of the 13 problems were addressed successfully except for those dealing with the problem of direct access.

As a result of this work we conclude that connectionist models are indeed extremely useful ways of characterizing the learning inflectional systems. We also conclude that the critiques of Pinker and Prince and Lachter and Bever erred in confusing conceptualizations with implementations.
REFERENCES


