Whatever Happened to the Past Tense Debate?

Steven Pinker
Harvard University

Twenty years ago, I began a collaboration with Alan Prince that has dominated the course of my research ever since. It began with a joint appearance in the MIT Center for Cognitive Science's seminar series, whose format faced off a speaker against two commentators. In May 1985, I had heard a presentation by James McClelland on a new model of language acquisition which challenged the assumptions of the approach to language – indeed, to the mind in general – in which I had been working, and which had been popular at MIT. I suggested that McClelland be invited to speak at the seminar, and asked Alan, who I had befriended when we were postdoctoral fellows at the Center five years earlier, if he would be willing to serve as the other commentator. I was familiar with the literature on language acquisition, having written a book on the topic (a book which McClelland had set up as his foil), but was unsure of my command of the relevant literature in phonology and morphology, which were among Alan's areas of expertise.

During our preparation for the event, Alan sent me a list of comments on the paper by McClelland and his collaborator, David Rumelhart, which left me flabbergasted. Not only had Alan identified some important flaws in their model, but pinpointed the rationale for the mechanisms that linguists and cognitive scientists had always taken for granted and that McClelland and Rumelhart were challenging – the armamentarium of lexical entries, structured representations, grammatical categories, symbol-manipulating rules, and modular organization that defined the symbol-manipulation approach to language and cognition. By pointing out the work that each of these assumptions did in explaining aspects of a single construction of language —the English past tense — Alan's talking points showed me the outline of a research program that could test the foundational assumptions of the dominant paradigm in cognitive science.

The MIT event itself was not quite fair to McClelland, who found himself double-teamed and on the wrong side of the home-field advantage. But everyone present agreed that our comments, which were some of the first focused criticisms of the new school of connectionism, should be written up. The following year (1986), Alan and I wrote a monster of a paper, which was widely circulated as a technical report and which embroiled us in a debate on connectionism, waged in part as a flame war in 1987-8 on the newly popular Arpanet. More constructively, it inspired a research program which led to a graduate course jointly taught at Brandeis and MIT, three grants, seven coauthored papers, a popular book (Pinker, 1999), and several jointly supervised graduate students (including John J. Kim, Greg Hickok, Gary Marcus, Michael Ullman, Sandeep Prasada, Fei Xu, William Snyder, Christ Collins, and Annie Senghas). Alan's departure of the Boston area for Rutgers in 1992, and our concurrent interests in other topics (Optimality Theory for Alan, evolutionary psychology and popular science writing for me) made direct collaboration more difficult to sustain, though the ideas laid out by Alan in his 1985 talking points continue to dominate my empirical research and shape my view of how the mind works.

The scientific story began when Rumelhart and McClelland (1986) took up the challenge to account for a phenomenon which had long served as the textbook case of people's ability to wield mental rules of grammar, the past tense suffix *-ed* which turns *walk* into *walked*. The past tense suffix was an apt choice because linguistic productivity had been demonstrated in children as young as four. When children are told, "Here is a man who knows how to *rick*; he did the same thing yesterday; he .....," they supply the appropriate novel form *ricked* (Berko, 1958). The standard explanation was that children were in command of the symbol-mainpulating rule "add *–ed*" (or some equivalent operation). Rumelhart and McClelland wanted to account for the phenomenon using an updated version of the theory of association by similarity, or stimulus generalization, which had been a centerpiece of associationist and behaviorist theories of the mind. A challenge for any theory of this phenomenon is that alongside the thousands of regular verbs that add *–ed,* there are about 180 irregular verbs of varying degrees of systematicity, such as *come-came, feel-felt*, and *teach-taught.* In the traditional theory, these were stored as pairs of words in memory, though there were also treatments, such as that of Chomsky and Halle (1968/1991), which generated them by minor rules.

Rumelhart and McClelland's classic associationist model of the past tense used the then-revolutionary, now-familiar approach of Parallel Distributed Processing or Connectionism. Their model acquired the past tense forms of hundreds of verbs, and that generalized properly to dozens of new verbs. More strikingly, it displayed a number of phenomena known to characterize children's behavior, most notably their overregularization of irregular verbs in errors such as *breaked* and *comed*. But the model had no explicit representation of words or rules; it simply mapped from units standing for the sounds of the verb stem to units standing for the sounds of the past tense form. Its apparent success led Rumelhart and McClelland to conclude that they had devised an alternative to a generative rule system which could account for the productivity seen when people produce and understand novel linguistic forms.

The commentary by Alan and me morphed into the technical report (an "MIT Center for Cognitive Science Occasional Paper") and then into an article (Pinker & Prince, 1988) published in a special issue of *Cognition* (Pinker & Mehler, 1988), accompanied by papers by Jerry Fodor and Zenon Pylyshyn (1988) and by Joel Lachter and Tom Bever (1988). Our criticisms did not go unanswered, nor did we let the answers to our criticisms go unanswered; our paper was the first of more than a hundred and fifty papers on the debate (see Marcus, 2000; Pinker, 1999; and Pinker & Ullman, 2002, for reviews). Every empirical claim in Pinker & Prince (1988) has been further examined, and twenty-five connectionist models purporting to fix the flaws of the RM model have been reported.

My graduate advisor Roger Brown once decried the lack of progress in much of psychology owing to the phenomenon in which "a large quantity of frequently conflicting theory and data can become cognitively ugly and so repellent as to be swiftly deserted, its issues unresolved." I like to think that the past-tense debate, now in its third decade, is a more hopeful case, despite the impression in some observers that it has reached a stalemate. In this paper I summarize my view of the current state of the art.

Not surprisingly, considering the longevity of the debate, the emerging picture embraces some elements of both sides (a resolution which Alan and I had

adumbrated in our 1988 paper). I think the evidence supports a modified version of the traditional words-plus-rules theory in which irregular forms, being unpredictable, are stored in memory as individual words, and regular forms are generated by rule, just like other productive complex constructions such as phrases and sentences. Memory, however, is not just a list of unrelated slots, but is partly associative: features are linked to features -- as in the connectionist pattern associators -- as well as words being linked to words. This means that irregular verbs are predicted to show the kinds of associative effects that are well-modeled by pattern associators: families of similar irregular verbs (e.g., *fling-flung, cling-clung, sling-slung*) are easier to store and recall (because similar verbs repeatedly strengthen a single set of connections for their overlapping material), and people are occasionally prone to generalize irregular patterns to new verbs similar to known ones displaying that pattern (e.g., as in *spling-splung*, because the new verbs contain features that have been associated with existing irregular families).

On the other hand, I believe the evidence shows that *regular* verbs are generated by a linguistic rule which concatenates a suffix to a stem – not an ad hoc operation specific to the past tense, but the result of a more general unification operation, which applies to the suffix *-ed* to satisfy multiple constraints imposed by the syntax, semantics, and morphology of a sentence (see Pinker & Jackendoff, 2005). Whereas irregular inflection is inherently linked to memorized words or forms similar to them, people can apply regular inflection to *any* word, regardless of its memory status. Many phenomena of linguistic structure and productivity can be parsimoniously explained by the simple prediction that whenever memorized forms are not accessed, for any reason, irregular inflection is suppressed and regular inflection is applied. Here I lay out the major phenomena and compare the explanations of the words-and-rules theory to those proposed by connectionist modelers in the two decades since my paper with Alan was published. In these models, both regular and irregular forms are generated by a single pattern associator memory; symbol concatenation operations and hierarchical linguistic structures are eschewed, as they were in the original RM model.

Of all the talking points in Alan's list, one of them – the qualitative difference between regular and irregular forms – generated the lion's share of research and commentary. But his other criticisms of the RM model (and, correspondingly, spotlights on the advantages of symbolic theories) are equally telling, and it is important to realize that we bent pretty far backwards in comparing pattern associator models of the past tense to the traditional words-and-rules theory as a general test case of the merits of connectionist versus traditional linguistic theories of language. Pattern associator models ignore so many key features of language that even if they did succeed in capturing the facts of the past tense, no one would be justified in concluding that they are viable models of language or have made rules obsolete, as many connectionists claim. Here are the rhetorical concessions we made in focusing the research on the regular-irregular distinction in the generation of English past tense forms, setting aside Alan's other points:

- First, the models have never seriously dealt with the problem of phonological representation, which in standard linguistic theories requires a hierarchical tree. Instead, they tend to use the problematic "Wickelfeature" representation (unordered sets of feature trigrams), or to artificially restrict

the vocabulary to a subset of English, such as CVC monosyllables.

- Second, most of the models account only for the *production* of past tense forms; they do not *recognize* such forms (e.g., for the purposes of speech production or grammaticality judgments), and therefore require a second, redundant network to do so. Clearly we need a representation of information for inflection that can be accessed in either direction, because people do not separately learn to produce and to comprehend the past tense form of every word.
- Third, the models are trained by a teacher who feeds them pairs consisting of a verb stem and its correct past tense form. This is based on the assumption that children, when hearing a past-tense form in their parents' speech, recognize that it *is* the past-tense form of a familiar verb, dredge the verb stem out of memory, feed it into their past tense network, and silently compare their network's output with what they just heard. How a child is supposed to do all this without the benefit of the lexical and grammatical machinery that the connectionists claim to have made obsolete has never been explained.
- Fourth, the models are studied in isolation of the rest of the language system. The modeler spoonfeeds verb stems and then peers at the model's output; the myriad problems of deciding whether to inflect a verb to start with, and if so with what inflection, are finessed. So as is the process of feeding the output into the right slot in a phrase or a larger word such as a compound.
- Fifth, the models are restricted to the relatively simple task of inflecting a single word. Complex, multi-affix morphology (as seen in polysynthetic languages), and all of syntax and compositional semantics, are almost entirely ignored.

In contrast, the words-and-rules theory treats the past tense as a mere example of the kind of symbol manipulation and modular design that characterizes the language system in general. It has already been scaled down from more articulated theories, and so does not face the severe problems of scaling up that would plague the pattern associator approach even if that approach succeeded at the past tense. That having been said, let's see whether they do succeed in five areas originally raised in "On Language and Connectionism" (Pinker & Prince, 1988).

1. Reliance of generalization on similarity

Alan and I noted that the RM model showed puzzling failures in generalizing the regular pattern to many novel verbs. For example, it turned *mail* into *membled*, and failed to generate any form at all for *jump* and *pump*. We conjectured that the problem came from the fact that the model generalizes by similarity to trained exemplars: new words overlap the phonological input units for previously trained similar words and can co-opt their connections to phonological output units to past tense sounds. It does not process *symbols* such as "Verb," which can embrace an entire class of words regardless of their phonological content. Therefore the model could not generate past-tense forms for simple verbs that were not sufficiently similar to those it had been trained on. Whereas irregular forms may indeed be

generalized by similarity, as in pattern associator models, the essence of regular generalizations is the ability to concatenate symbols.

Sandeep Prasada and I (1993) (see also Egedi & Sproat, 1991; Sproat, 1992) confirmed the conjecture by showing that the trained RM model did a reasonably good impersonation of the human being when it comes to generalizing irregular patterns: they both converted *spling* to *splung*, generalizing the pattern from similar *cling-clung.* But with the regular words, people and the model diverged: both people and the model could convert *plip* (similar to existing verbs such as *flip* and *clip)* to *plipped*, but only people, not the model, could convert *ploamph* (not similar to any existing verb) to *ploamphed.* The model instead produced gibberish such as *ploamph-bro*, *smeej-leafloag,* and *frilg-freezled.* Lacking a symbol, and confined to associating bits of sound with bits of sound, the model has nothing to fall back on if a new item doesn't overlap similar, previously trained items, and can only cough up a hairball of the bits and pieces that are closest to the ones that it *has* been trained on. People, in contrast, reason that a verb is a verb, and, no matter how strange the verb sounds, they can hang an *-ed* on the end of it.

The problem of computing coherent past tense forms for novel-sounding verbs still has no satisfactory solution in the framework of standard connectionist pattern-associator memories. (See Marcus, 2000, for extensive analysis, as well as Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995, and Prasada & Pinker, 1993). Several modelers, stymied by the models' habit of outputting gibberish, have hardwired various patches into their model that are tailor-made for regular verbs. One team of modelers included a second pathway of connections that linked every input unit to its twin in the output, implementing by brute force the copying operation of a rule (MacWhinney & Leinbach, 1991). Another team added an innate clean-up network in which the units for *-ed* strengthen the units for an unchanged stem vowel and inhibit the units for a changed vowel, shamelessly wiring in the English past tense rule (Hare, Elman, & Daugherty, 1995). And many connectionist modelers have given up on trying to generate past-tense forms altogether. Their output layer contains exactly one unit for every past tense suffix or vowel change, turning inflection into a multiple-choice test among a few innate possibilities (e.g., Hare & Elman, 1992; Nakisa & Hahn, 1996). To turn the choice into an actual past-tense form, some *other* mechanism, hidden in the wings, would have to copy over the stem, find the pattern corresponding to the chosen unit, and apply the pattern to the stem. That mechanism, of course, is called a rule, just what connectionists claim to be doing without.

2. Systematic regularization

Alan pointed out that some irregular verbs mysteriously show up in regular garb in certain contexts. For example, you might say *All my daughter's friends are low-lifes*, not *low-lives*, even though the ordinary irregular plural of *life* is *lives*. People say *Powell ringed the city with artillery*, not *rang*, and that a politician *grandstanded*, not *grandstood*. This immediately shows that sound alone cannot be the input to the inflection system, because a given input, say *life*, can come out the other end of the device either as *lifes* or as *lives*, depending on something else.

What is that something else? Connectionists have repeatedly suggested that it is meaning: a semantic stretching of a word dilutes the associations to its irregular past

tense form, causing people to switch to the regular (e.g., Harris, 1992; Lakoff, 1987; MacWhinney & Leinbach, 1991). But that this is just false. In the vast majority of cases in which an irregular word's meaning changes, the irregular form is unchanged. For example, if you use a noun metaphorically, that the irregular plural is untouched: *straw men, snowmen, sawteeth, God's children* (not *mans, tooths,* or *childs).* And English has hundreds of idioms in which a verb takes on a wildly different meaning, but in all cases it keeps its irregular past tense form: *cut a deal* (not *cutted*), *took a leak, caught a cold*, *hit the fan*, *blew them off, put them down, came off well, went nuts*, and countless others (Kim, Marcus, Pinker, Hollander, & Coppola, 1994; Kim, Pinker, Prince, & Prasada, 1991; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995; Pinker & Prince, 1988). So it is not enough simply to add a few units for meaning to an associative memory and hope that any stretch of meaning will cut loose an irregular form and thereby explain why people say *low-lifes* and *grandstanded.*

Equally unsatisfactory is the suggestion that people regularize words to avoid ambiguity and make themselves clear (Daugherty, MacDonald, Petersen, & Seidenberg, 1993; Harris, 1992; Shirai, 1997). Many idioms are ambiguous between literal and idiomatic senses, such as *bought the farm* and *threw it up*, and some are ambiguous with other idioms as well: *blew away*, for example, could mean "wafted," "impressed," or "assassinated"; *put him down* could mean "lower," "insult," or "euthanize." But that doesn't tempt anyone to single out one of the meanings in each set by saying *buyed the farm, throwed up, blowed him away*, or *putted him down*. Conversely, the past tense of to *grandstand* is *grandstanded*, not *grandstood*, but *grandstood* would be perfectly unambiguous if anyone said it. The same is true of *Mickey Mice, high-stuck* and *lowlives*, which would be perfectly clear, especially in context. But with these unambiguous words people are tempted, even compelled, to use a regular past-tense form.

A better theory (Kiparsky, 1982; Selkirk, 1982; Williams, 1981) says that *headless* words become regular (Kim, Marcus, Pinker, Hollander, & Coppola, 1994; Kim, Pinker, Prince, & Prasada, 1991; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995; Pinker, 1999; Pinker & Prince, 1988). The point of rules of grammar is to assemble words in such a way that one can predict the properties of the new combination from the properties of the parts and the way they are arranged. That is true not just when we string words into sentences, but when we string bits of words into complex words. Start with the noun *man*. Combine it with *work*, to produce a new word, *workman*. The scheme for deducing the properties of the new word from its parts is called the right hand head rule: take the properties of the rightmost element and copy them up to apply to the whole word. What kind of word is *workman*? It's a noun, because *man,* the rightmost element, is a noun, and the nounhood gets copied up to apply to the whole new word. What does *workman* mean? It's a kind of man, a man who does work: the *meaning* of *man* is passed upstairs. And what is the plural of *workman*? It's *workmen*, because the plural of *man* is *men*, and that information, too, gets copied upstairs too.

But there is a family of exceptions: headless words, which don't get their features from the rightmost morpheme. In some compound words, for example, the meaning pertains to something that the rightmost noun *has* rather than something the rightmost noun *is*. For example, what is a low-life? A kind of life? No, it is a kind of

person, namely, a person who has (or leads) a low life. In forming the word, you have to turn off the right hand head rule – that is, plug the information pipeline from the root in memory to the whole word -- in order to prevent the word from meaning a kind of life. If the pipeline is plugged, there is no longer any way for the irregular plural of *life, lives*, to percolate up. That information is sealed in memory, and the regular "add *-s*" rule steps in as the default. Other examples include *still-lifes* (not *still-lives*), is not a kind of life but a kind of painting, and *sabretooths*, not *sabre-teeth*, because the word refers not to a kind of tooth but to a kind of cat.

Another example showing off this mental machinery comes from verbs that are based on nouns. We say that the artillery *ringed the city*, not *rang*, because the verb comes from a noun: *to ring* in this sense means to form a ring around. To get a noun to turn into a verb, the usual percolation pipeline has to be blocked, because ordinarily the pipeline allows part-of-speech information be copied from the root to the newly formed word. And that blocked pipeline prevents any irregularity associated with the sound of the verb from applying to the newly formed word. For similar reasons, we say that a politician *grandstanded*, not *grandstood*, because the verb comes from the noun play to the grandstand. Note, by the way, that no machinery has been posited specifically to generate the regularizations; the right-hand-head rule is the standard mechanism to account for morphological composition in general. A single mechanism accounts for morphological composition, for regularizations caused by headless compounds such as *low-lifes,* for regularizations caused by denominal verbs such as *ringed*, and for some half-dozen other grammatical quirks (Marcus et al., 1995; Pinker, 1999; Pinker & Ullman, 2002).

How can a connectionist model account for these facts? Daugherty, MacDonald, Peterson, and Seidenberg (1993) added input nodes representing the degree of semantic distance of the verb from a homophonous noun. From there it is trivial to train the network to have these nodes turn off irregular patterns and turn on the regular one. But these strange nodes are not part of the semantic representation of a verb itself, but an explicit encoding of the verb's relation to the noun that heads it -- that is, a crude implementation of morphological structure, wired in to duplicate phenomena that had been discovered and explained by the linguistic structure account. Daugherty et al. tried to motivate the representation with reference to a suggestion by Harris (1992) that speakers regularize denominals to enhance communication (presumably to disambiguate homophones), but as I have pointed out, the evidence runs against hypothesis: there are hundreds of pairs of ambiguous verbs with irregular verb roots (*blew away* = "wafted; assassinated; impressed"), and they do not regularize, and the vast majority of verbs with noun roots are not ambiguous (e.g., *grandstanded*), and they do regularize. A final problem is that Daugherty et al. had to train their model on regular past tenses of denominal verbs homophonous with irregulars (about 5% of the training exemplars). But such verbs, though scientifically interesting test cases, are used extremely rarely, and speakers cannot depend on having heard them regularized (Kim, et al., 1994).

3. Childhood overregularization errors

Children frequently make errors such as *We holded the baby rabbits* and *The alligator goed kerplunk* (Cazden, 1968; Ervin & Miller, 1963). The words-and-rules theory offers a simple explanation: children's memory retrieval is less reliable than

adults'. Since children haven't heard *held* and *came* and *went* very often (because they haven't lived as long), they have a weak memory trace for those forms. Retrieval will be less reliable, and as long as the child has acquired the regular rule, he or she will fill the vacuum by applying the rule, resulting in an error like *comed* or *holded* (Marcus, et al., 1992).

Evidence that weak memory is a factor comes from many sources, summarized in Marcus et al. (1992). For example, Marcus and I found that the more often a child's parent uses an irregular when talking to the child, the less often the child makes an error on it. The theory explains why children, for many months, produce no errors with these forms – at first they say *held* and *came* and *went*, never *holded* and *comed* and *goed*. We proposed that that is the point at which the child has just acquired the *-ed* rule. Very young children say things like *Yesterday we walk,* leaving out past tense marking altogether. They pass from a stage of leaving out the *ed* more often than supplying it to a stage of supplying it more often than leaving it out, and the transition is exactly at the point in which the first error like *holded* occurs. This is what we would expect if the child has just figured out that the past tense rule in English is add *-ed*. Before that, if the child failed to come up with an irregular form, he had no choice but to use it in the infinitive: *Yesterday, he bring*...; once he has the rule, he can now fill the gap by over-applying the regular rule, resulting in *bringed*.

In contrast, the connectionist accounts of the transition are incompatible with many of the facts. For starters, the basic assumptions of Rumelhart and McClelland's developmental account – that vocabulary growth leads to an increase in the proportion of regular verbs fed into the network – is deeply problematic. Children presumably learn as they listen to the speech coming out of the mouths of their parents, not by scanning their own mental dictionaries and feeding each verb into their network once per pass. That implies that it should be the percentage of *tokens*, not the percentage of *types* (vocabulary items), that must be counted. And the percentage of tokens that are regular remains constant throughout development, because irregular verbs are so high in token frequency that they remain predominant even as the number of regular types increases.

The percentage of regular types does increase as the child's vocabulary expands, of course, because there is a fixed number of irregular verbs in the language and the child will eventually learn them all and thereafter expand his vocabulary only by learning regular forms. But the rate of increase in regular vocabulary is *negatively*, not positively, correlated with overregularization in children's speech over time. That is because children's vocabulary spurt, which Rumelhart and McClelland credited for the onset of overregularization, occurs a full year before the first overregularization errors. Plunkett and Marchman (Plunkett & Marchman, 1991; Plunkett & Marchman, 1993) claimed to have devised a new PDP model that began to overregularize like children without an unrealistic change in the mixture of regular and irregular verbs in the input (see also Marchman & Bates, 1994), but Marcus et al. (1992) and Marcus (1995) have shown that this claim is belied by their own data, and that the developmental curves of the Plunkett-Marchman models are qualitatively different from those of children in several ways.

4. Neuropsychological dissociations

A particularly direct test of the words-and-rules theory consists of cases in which the human memory system is directly compromised by neurological damage or disease. Ullman and I (Ullman, et al., 1997) asked a variety of neurological patients to fill in the blank in items like "Everyday I like to (verb); yesterday, I ..." We tested patients with anomia, an impairment in word finding, often associated with damage to the posterior perisylvian region of the left hemisphere; such patients can often produce fluent and mostly grammatical speech, suggesting that their mental dictionaries are more impaired than their mental grammars. With such patients, we found that irregular verbs are harder than regulars, which fits the theory that irregulars depend on memory whereas regulars depend on grammar. We also predicted and observed regularization errors like *swimmed*, which occur for the same reason that children (who also have weaker memory traces) produce such errors: they cannot retrieve *held* from memory in time. And the patients are relatively unimpaired in doing a *wug*-test (*Today I wug, yesterday I wugged*), because that depends on grammar, which is relatively intact.

Conversely, brain-injured patients with agrammatism (a deficit in stringing words together into grammatical sequences, often associated with damage to anterior perisylvian regions of the left hemisphere) should show the opposite pattern: they should have more trouble with regulars, which depend on grammatical combination, than with irregulars, which depend on memory. They should produce few errors like *swimmed*, and they should have trouble doing the *wug*-test. And that is exactly what happens (Marin, Saffran, & Schwartz, 1976; Ullman, et al., 1997)). These dissociations are part of a growing set of neuropsychological studies showing that the processing of regular forms and the processing of irregular forms take place in different sets of brain areas (Pinker, 1997; 1999; Pinker & Ullman, 2002; Ullman et al., 2005).

Double dissociations are difficult to explain in uniform pattern associators, because except for artificially small networks, "lesioning" the networks hurts the irregular forms more than regular ones (Bullinaria & Chater, 1995). One exception is a simulation of the past tense by Marchman (1993), which seemed to go the other way. But 60 percent of Marchman's "irregular" items were no-change verbs such as *hit-hit*, which use a highly predictable and uniform mapping shared with the regular verbs. This artificial word list, and the fact that the model didn't do well with the regular verbs even *before* it was lesioned, explain the anomalous result. A more recent model by Joanisse & Seidenberg (1999) conceded that distinct subsystems have to be lesioned to produce double dissociations. Although they called these modules "phonologica" and "semantic," the semantic module was in fact a lexicon: it had one unit dedicated to each word, with no representation of meaning. The finding that lesioning a lexicon differentially impairs irregular inflection is exactly what the words-and-rules model predicts. Moreover, the model failed to duplicate the finding that agrammatic patients have more trouble with regular than irregular verbs (Ullman, et al., 1997, 2005). Lesioning the phonology module caused a consistent selective deficit only with novel verbs; regulars were no harder than irregulars. The report also claims that because a novel form has no meaning, "the only way to generate its past tense is by analogy to known phonological forms" (Jonaisse & Seidenberg 1999: 81). This predicts that patient groups should have parallel

tendencies to generalize regular and irregular inflection to novel words (plammed and splung, respectively), whereas in fact these tendencies dissociate (Ullman et al., 2005). Finally, the model predicts that selective difficulty with irregular forms should depend on semantic deficits. Michele Miozzo (2003) reports an anomic patient who had difficulty accessing word forms but not word meanings; nonetheless, he had trouble with irregulars but not with regulars.

5. Crosslinguistic comparisons

Many connectionists have tried to explain away our findings on English by pointing to a possible confound, type frequency: regular verbs are the majority in English. Only about 180 verbs in modern English are irregular, alongside several thousand regular verbs. Since pattern associators generalize the majority pattern most strongly, it is conceivable that a pattern associator that was suitably augmented to handle grammatical structure would have the regular pattern strongly reinforced by the many regular verbs in the input, and would come to generalize it most strongly.

Taking this argument seriously requires yet another act of charity. Pattern associators are driven by tokens rather than types: the models are said to be learn in response to actual utterances of verbs, in numbers reflecting their frequencies of usage, rather than in response to vocabulary entries, inputted once for each verb regardless of its frequency of usage. So differences in the sheer number of vocabulary items in a language should not have a dramatic effect, because the irregular forms are high in token frequency and dominate tokens of speech (with different numbers of regular forms rotating in and out of a minority of the conversational slots). Moreover, no pattern associator model yet proposed has plausibly handled the various grammatical circumstances involving headlessness (*low-lifes*, *ringed the city,* and so on) in which irregular forms systematically regularize.

But many connectionist researchers have held out the greater type frequency of regular verbs in English as the main loophole by which future pattern associators might account for the psycholinguistic facts reviewed herein (Bybee, 1995; MacWhinney & Leinbach, 1991; Seidenberg, 1992; see Marcus et al., 1995, for quotations). To seal the case for the word-rule theory it would be ideal to find a language in which the regular (default) rule applies to a minority of forms in the language. Note that some connectionists, reasoning circularly, treat this prediction as an oxymoron, because they *define* regular as pertaining to the most frequent inflectional form in a language and irregular to pertain to the less frequent forms. But we are considering a psycholinguistic definition of regular as the default operation produced by a rule of grammatical composition and irregular as a form that must be specially stored in memory; the number of words of each kind in the language plays no part in this definition.

One language that displays this profile is German (Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995). The plural comes in eight forms: four plural suffixes (*-e, -er, -en, -s*, and no suffix), some of which can co-occur with an altered (umlauted) stem vowel. The form that acts most clearly as the default, analogous to English *-s*, is *-s*. German allows us to dissociate grammatical regularity from type frequency (see Marcus et al., 1995, for a far more extensive analysis). In English, *-s* is applied to more than 99% of all nouns; in German, *-s* is applied to only about 7% of nouns.

Despite this enormous difference, the two suffixes behave similarly across different circumstances of generalization. For example, in both languages, the *-s* suffix is applied to unusual-sounding nouns (*ploamphs* in English, *Plaupfs* in German), to names that are homophonous with irregular nouns (*the Julia Childs*, *die Thomas Manns*), and to many other cases of systematic regularization. Moreover, German-speaking children frequently overregularize the suffix in errors such as *Manns*, analogous to English-speaking children's *mans*. So despite the relatively few nouns in German speech taking an *-s*-plural, it shows all the hallmarks of a rule product, showing that the signs of a rule cannot be explained by sheer numbers of vocabulary items.

There is one final escape hatch for the connectionist theory that the generalizability of regular patterns comes from the statistics of regular words in a language. Several connectionist modelers have replied to our arguments about German by saying that it may not be the *number* of regular words that is critical so much as the *scattering* of regular words in phonological space (Forrester & Plunkett, 1994; Hare, Elman, & Daugherty, 1995; Nakisa & Hahn, 1996; Plunkett & Nakisa, 1997). Suppose irregulars fall into tight clusters of similar forms (*sing, ring, spring; grow, throw, blow,* and so on), while regulars are kept out of those clusters but are sprinkled lightly and evenly throughout no-man's-land (*rhumba'd, out-Gorbachev'd, oinked*, and so on). Then one can design pattern associators that devote some of their units and connections to the no-man's-land, and they will deal properly with any subsequent strange-sounding word. These models cannot be taken seriously as theories of a human child, because they have the inflections of a language innately wired in, one output node per inflection, and merely learn to select from among them. And as usual, the problem of rootless and headless words is ignored. But bending over backwards even further, we can test the general idea that certain patterns of clustering among regular and irregular sounds are necessary for people to generalize the regular inflection freely.

In any case, Iris Berent has nailed that escape hatch shut (Berent, Pinker, & Shimron, 1999). In Hebrew, regular and irregular nouns live cheek-by-jowl in the same phonological neighborhoods. Irregular nouns do not carve out their own distinctive sounds, as they do in English. Nonetheless, the regular plural suffixes *–im* and *–ot* behave similarly to *–s* in English and German: speakers apply them to unusual sounding nouns, and to names based on irregular nouns (analogous to our *The Childs*, *The Manns*). Moreover, using other examples from Hebrew which unconfound variables that are confounded in English, Berent and I have shown (Berent, Pinker, & Shimron, 2002) that the greater preservation of the stem in English regular as opposed to irregular verbs (an example of faithfulness in Optimality-Theoretic terms) also cannot account for the psychological differences between regular and irregular forms in speakers' judgments.

6. Summary of Empirical Comparisons Between Connectionist and Symbolic Theories of the Past Tense

The preceding comparisons have shown that despite the identical function of regular and irregular inflection, irregulars are avoided, but the regular suffix is applied freely in a variety of circumstances, from *chided* to *ploamphed* to *lowlifes* to *anomia*, that have nothing in common except a failure of access to information in

memory. Crucially, I have presented these diverse cases precisely because they are so heterogeneous and exotic. Even if a separate connectionist model were devised that successfully accounted for each of these phenomena in a psychologically plausible way – and that is far from the case – it would be hard to treat the set of models as a psychologically plausible theory of language. Clearly we don't have separate innate neural mechanisms each designed to generate regular forms in one of these cases. Rather, the repeated appearance of the regular pattern falls out of the simple theory that the rule steps in whenever memory fails, regardless of the reason that memory fails. And that in turn implies that rules and memory are different systems.

Let me mention the remaining arguments for the connectionist approach to the past tense and related phenomena. Occasionally I am asked whether it might be unparsimonious to posit two mechanisms, rather than trying to handle all the phenomena in one. (In a characteristically erudite allusion, Alan characterized this sentiment with a line from a poem by Empson: "Lucretius could not credit centaurs / Such bicycle he deemed asynchronous.") In general I am unmoved by the a priori argument that the human mind, or any of its major subsystems, ought to contain exactly one part. In the case of inflection in particular, parsimony works in favor of the words-and-rules theory. No one (including the connectionists) has ever tried to model simple words and productive sentences in a single mechanism (that is, uninflected monomorphemes, and novel sentences assembled on the fly). Even in the connectionist literature, the models of the lexicon and the models of sentence processing (e.g., the recurrent network of Elman, 1990) are distinct, and even Rumelhart & McClelland and MacWhinney & Leinbach admit in the fine print that a more realistic model than theirs would need a separate lexicon. If one has a mechanism for storing and retrieving words, and one has a separate mechanism for assembling and parsing sentences, one already has words and rules, exactly the mechanisms needed to handle irregular and regular forms. In other words the word/rule distinction is needed for language in general; it was not invented to explain regular and irregular forms per se.

According to a second objection, linguists themselves have shown that the distinction between words and rules is obsolete. After all, there are novel complex words like *unmicrowaveability* that have to be generated by rules, and there are phrasal idioms such as *hit the fan* and *beat around the bush* that have to be memorized like words (di Sciullo & Williams, 1987; Jackendoff, 1997). This argument merely underscores the fact that the word *word* is highly ambiguous, with at least four senses. In the sense of intended by the words-and-rules theory, "word" refers to what Di Sciullo and Williams call listemes: language chunks of any size, from morpheme to proverb, that are not fully compositional and therefore have to be memorized. And "rule" is not intended to refer narrowly to a classic rewrite production such as S → NP VP; it is meant to refer more broadly to any productive, combinatorial operations on symbolic structures, including principles, constraints, unification, optimality, and so on. A more accurate (but less euphonious) title for the theory would have been *Listemes and Combinatorial Symbolic Operations* (see also Pinker & Jackendoff, 2005).

A third objection is that Alan and I were beating a dead horse by testing the old Rumelhart-McClelland model. All the connectionists agree that it was a simplistic

first attempt; there are new models that do much better and account for each of the phenomena that the RM model failed on.

In fact, twenty years and twenty-five models later, the RM model may still be the best connectionist model of the past tense. For one thing, its supposedly low-tech features – the lack of a hidden layer, and its Wickelfeature representation of the input and output – turn out to make little or no difference. Richard Sproat and Dana Egedi (Egedi & Sproat, 1991) did head-to-head comparisons of the original model and a version with a hidden layer and a state-of-the-art representation and output decoder. The souped-up version had the same problems as the original.

Many connectionist fans are surprised to learn that the RM model isn't any worse than its successors, because the standard doctrine in connectionist modeling is that hidden-layer models are more powerful than perceptrons. But as Marvin Minsky and Seymour Papert (1988) have pointed out, one can compensate for the lack of a hidden layer by beefing up the input representation, and that's what Rumelhart and McClelland, perhaps inadvertently, did. Every word got a "blurred" input representation, in which a smattering of incorrect units were activated for that word together with the correct ones. The blurring was not, however, random noise: the same set of incorrect units got activated for a given word every time it was fed into the network. Moreover, many of the incorrect units code for sound sequences that cannot exist in English words. Thus each set of "blurred" units can serve as a unique code for that lexical item. This was particularly effective because many of the blurred units represented Wickelfeatures that are phonologically impossible, so those units didn't have to do any work in the sound-to-sound mapping, and were therefore available to code individual lexical entries without interference. This compensated nicely for the lack of a hidden layer, which under ordinary training circumstances comes to code indirectly for distinct lexical entries. Once again, traditional linguistic notions (in this case a lexical entry) have to be reintroduced into the pattern associator models through the back door.

But most important, the RM network remains the only model with empirical content – that is, the only model whose behavior makes a novel and correct prediction about human behavior. Rumelhart and McClelland built a model to compute and generalize the past tense, and the model not only did that, but successfully predicted what kinds of irregular forms children find more or less easy, and it also successfully predicted several forms of such errors. (The model made some incorrect predictions as well, for interesting reasons.) That kind of predictive record can't be ignored, which is why Alan Prince and I noted back in 1988 that the model, at the very least, had to be taken seriously as capturing something about the memory system in which irregular forms are stored.

In contrast, the immediate follow-up models either made empirical predictions that are demonstrably false (see Marcus, 1995, 2000; Kim et al., 1994; Marcus et al., 1995; Berent, Pinker, & Shimron, 1999, 2002), or didn't make predictions at all, because they were kluged by hand to mimic a specific phenomenon that our group had previously documented. The modelers seem content to show that some connectionist model or other can mechanically generate some behavior (true enough, in the same sense that some Fortran program or other can generate the behavior), without showing that the model is true or even remotely plausible (see Pinker & Ullman, 2003.)

One irony of the past-tense debate is that it's often framed as a nature-nurture battle, with the connectionists on the side of nurture. But in fact the connectionist models build in innate features specific to the English past tense that would make Jerry Fodor blush – such as a layer of nodes whose only purpose is to generate anomalies such as *ringed the city* and *grandstanded*, or an output layer that consists of exactly one innate node for each inflection in English, with the model merely selecting among them in a multiple-choice task. Indeed, those innate-inflection networks have taken over in the most recent generation of connectionist models of inflection (by Elman, Nakisa & Hahn, and others); Rumelhart & McClelland's more ambitious goal of c*omputing* the output form has largely been abandoned without comment. As mentioned earlier, there is a double irony in these models: the English inflections are innate, and since the model only selects the appropriate suffix or vowel-change, some unmentioned postprocessor has to apply the suffix or vowel-change to the stem to generate the actual form. That postprocessor, of course, is what linguists call a rule – exactly what the models are supposed to be doing without.

Early in the debate, Alan and I were often asked if we would deny that *any* connectionist model could *ever* handle inflection. Of course not! We were skeptical only of the claim that the current favored style – a single pattern-associator –can handle it in a psychologically realistic way. A neural network model consisting of an associative memory (for words, including irregulars) and a hierarchical concatenator (for combinatorial grammar, including regulars) could (if the details were done correctly) handle all the phenomena at issue. Our objections are aimed not at connectionism, but at the fashion of denying compositional structure and shoehorning phenomena into a single uniform net. We were open to the possibility that that take structure seriously, such as the various proposals by Alan's collaborator Paul Smolensky (1990; Smolensky & Legendre, 2006), and by Lokendra Shastri (Shastri, 1999; Shastri & Ajjanagadde, 1993), John Hummel and colleagues (1992; 1997), and others.

7. A final word

Anyone who has spent a career in academia will have met many extraordinary minds. Yet even against this background Alan stands out. His scientific and literary erudition, analytical horsepower, lightning wit, intellectual generosity, and insight into the nature of language are without parallel among scholars in the human sciences. His collaboration and friendship have led to the scientific work I am most proud of and to countless moments of intellectual illumination and sheer fun. It's an honor to recognize his accomplishments and to wish him a happy sixtieth birthday.

References

Berent, I., Pinker, S., & Shimron, J. (1999). Default nominal inflection in Hebrew: Evidence for mental variables. *Cognition, 72*, 1-44.

Berent, I., Pinker, S., & Shimron, J. (2002) The nature of regularity and irregularity: Evidence from Hebrew nominal inflection. *Journal of Psycholinguistic Research, 31(5),* 459-502

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

Bullinaria, J.A., & Chater, N. (1995). Connectionist modeling: Implications for cognitive neuropsychology. *Language and Cognitive Processes, 10*, 227-264.

Bybee, J.L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*, 425-455.

Cazden, C.B. (1968). The acquisition of noun and verb inflections. *Child Development, 39*, 433-448.

Chomsky, N., & Halle, M. (1968/1991). *The sound pattern of English*. Cambridge, MA: MIT Press.

Daugherty, K.G., MacDonald, M.C., Petersen, A.S., & Seidenberg, M.S. (1993). Why no mere mortal has ever flown out to center field but people often say they do, *Fifteenth Annual Conference of the Cognitive Science Society*: Erlbaum.

di Sciullo, A.M., & Williams, E. (1987). *On the definition of word.* Cambridge, MA: MIT Press.

Egedi, D.M., & Sproat, R.W. (1991). Connectionist networks and natural language morphology. Murray Hill, NJ: Linguistics Research Department, Lucent Technologies.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Ervin, S.M., & Miller, W. (1963). Language development. In H.W. Stevenson (Ed.), *Child Psychology: The sixty-second yearbook of the National Society for the Study of Education, Part I.* Chicago: University of Chicago Press.

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

Forrester, N., & Plunkett, K. (1994). Learning the Arabic plural: The case for minority default mappings in connectionist networks. A. Ram & K. Eiselt (Eds.), *Sixteenth Annual Conference of the Cognitive Science Society*, Mahwah, NJ: Erlbaum.

Hare, M., & Elman, J. (1992). A connectionist account of English inflectional morphology: Evidence from language change, *Fourteenth Annual Conference of the Cognitive Science Society*: Erlbaum.

Hare, M., Elman, J., & Daugherty, K. (1995). Default generalization in connectionist networks. *Language and Cognitive Processes, 10*, 601-630.

Harris, C.L. (1992). Understanding English past-tense formation: The shared meaning hypothesis, *Fourteenth Annual Conference of the Cognitive Science Society*: Erlbaum.

Hummel, J.E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*, 480-517.

Hummel, J.E., & Holyoak, K.J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 427- 466.

Jackendoff, R. (1997). *The architecture of the language faculty.* Cambridge, MA: MIT Press.

Joanisse, M.F., & Seidenberg, M.S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Science, USA, 96*, 7592-7597.

Kim, J.J., Marcus, G.F., Pinker, S., Hollander, M., & Coppola, M. (1994). Sensitivity of children's inflection to morphological structure. *Journal of Child Language, 21*, 173-209.

Kim, J.J., Pinker, S., Prince, A., & Prasada, S. (1991). Why no mere mortal has ever flown out to center field. *Cognitive Science, 15*, 173-218.

Kiparsky, P. (1982). Lexical phonology and morphology. In I.S. Yang (Ed.), *Linguistics in the morning calm, Vol. 3-91*. Seoul: Hansin.

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.

Lakoff, G. (1987). Connectionist explanations in linguistics: Some thoughts on recent anti-connectionist papers. Posting in connectionists@cs.cmu.edu.

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

Marchman, V. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience, 5*, 215-234.

Marchman, V.A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21(2)*, 339-366.

Marcus, G.F. (1995). The acquisition of inflection in children and multilayered connectionist networks. *Cognition, 56*, 271-279.

Marcus, G.F. (2000). *The algebraic mind: Reflections on connectionism and cognitive science.* Cambridge, MA: MIT Press.

Marcus, G.F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology, 29*, 189-256.

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*.

Marin, O., Saffran, E.M., & Schwartz, M.F. (1976). Dissociations of language in aphasia: Implications for normal function. In S.R. Harnad, H.S. Steklis, & J. Lancaster (Eds.), *Origin and evolution of language and speech, Vol. New York Academy of Sciences*. Annals of the New York Academy of Sciences. Vol. 280.: New York.

Minsky, M., & Papert, S. (1988). Epilogue: The new connectionism. In Minsky & Papert (Eds.), *Perceptrons: Expanded Edition*. Cambridge, MA: MIT Press.

Miozzo, M. (2003). On the processing of regular and irregular forms of verbs and nouns: Evidence from neuropsychology. *Cognition*, *87*, 101-127.

Nakisa, R.C., & Hahn, U. (1996). Where defaults don't help: The case of the German plural system. G.W. Cottrell (Ed.), *Eighteenth Annual Conference of the Cognitive Science Society*: Erlbaum.

Pinker (1997). Words and rules in the human brain. *Nature, 387*, 547-548.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. New York: Basic Books.

Pinker, S., & Mehler, J. (1988). Connections and symbols. Cambridge, MA: MIT Press.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition, 95*, 201-236.

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

Pinker, S. & Ullman, M. (2002) The past and future of the past tense. *Trends in Cognitive Science, 6*, 456-463.

Pinker, S. & Ullman, M. (2003) Beyond one model per phenomenon. *Trends in Cognitive Science*, 7, 108-109.

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. (1993). From rote learning to system building. *Cognition, 48*, 21-69.

Plunkett, K., & Nakisa, R. (1997). A connectionist model of the Arab plural system. *Language and Cognitive Processes, 12*, 807-836.

Prasada, S., & Pinker, S. (1993). Generalizations of regular and irregular morphological patterns. *Language and Cognitive Processes, 8*, 1-56.

Rumelhart, D.E., & McClelland, J.L. (1986). On learning the past tenses of English verbs. Implicit rules or parallel distributed processing? In J.L. McClelland, D.E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 2: Psychological and biological models.* Cambridge, MA: MIT Press.

Seidenberg, M. (1992). Connectionism without tears. In S. Davis (Ed.), *Connectionism: Theory and practice*. New York: Oxford University Press.

Selkirk, E.O. (1982). *The syntax of words.* Cambridge, MA: MIT Press.

Shastri, L. (1999). Advances in SHRUTI: A neurally motivated model of relational knowledge representation and rapid inference using temporal synchrony. *Applied Intelligence, 11*, 79-108.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences, 16*, 417-494.

Shirai, Y. (1997). Is regularization determined by semantics, or grammar, or both? Comments on Kim, Marcus, Pinker, Hollander & Coppola. *Journal of Child Language, 24*, 495-501.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence, 46*, 159-216.

Smolensky, P., & Legendre, G. (2006) *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar.* Volume II*: Linguistic and Philosophical Implications*. Cambridge, MA: MIT Press.

Sproat, R. (1992). *Morphology and computation*. Cambridge, MA: MIT Press.

Ullman, M., Corkin, S., Coppola, M., Hickok, G., Growdon, J.H., Koroshetz, W.J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience, 9*, 289-299.

Ullman, M. T., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language, 93*(2), 185-238.

Williams, E. (1981). On the notions "lexically related" and "head of a word.". *Linguistic Inquiry, 12*, 245-274.

Department of Psychology
Harvard University
William James Hall 970
Cambridge, MA 02138
pinker@wjh.harvard.edu
http://pinker.wjh.harvard.edu/