The Past-Tense Debate

The past and future of the past tense

Steven Pinker and Michael T. Ullman

What is the interaction between storage and computation in language processing? What is the psychological status of grammatical rules? What are the relative strengths of connectionist and symbolic models of cognition? How are the components of language implemented in the brain? The English past tense has served as an arena for debates on these issues. We defend the theory that irregular past-tense forms are stored in the lexicon, a division of declarative memory, whereas regular forms can be computed by a concatenation rule, which requires the procedural system. Irregulars have the psychological, linguistic and neuropsychological signatures of lexical memory, whereas regulars often have the signatures of grammatical processing. Furthermore, because regular inflection is rule-driven, speakers can apply it whenever memory fails.

For fifteen years, the English past tense has been the subject of a debate on the nature of language processing. The debate began with the report of a connectionist model by Rumelhart and McClelland [1] and a critique by Pinker and Prince [2], and has since been the subject of many papers, conferences and simulation models [3–7] (see also McClelland and Patterson in this issue [8]).

The past tense is of theoretical interest because it embraces two strikingly different phenomena. Regular inflection, as in walk-walked and play-played, applies predictably to thousands of verbs and is productively generalized to neologisms such as spam-spanmed and mosh-moshed, even by preschool children [9]. Irregular inflection, as in come-came and feel-felt, applies in unpredictable ways to some 180 verbs, and is seldom generalized; rather, the regular suffix is often overgeneralized by children to these irregular forms, as in held-held and broke-broke [10,11]. A simple explanation is that irregular forms must be stored in memory, whereas regular forms can be generated by a rule that suffices -ed to the stem [12,13]. Rumelhart and McClelland challenged that explanation with a pattern-associate model (RMM) that learned to associate phonological features of the stem with phonological features of the past-tense form. It thereby acquired several hundred regular and irregular forms and overgeneralized -ed to some of the irregulars.

The past tense has served as one of the main empirical phenomena used to contrast the strengths and weaknesses of connectionist and rule-based models of language and cognition [8]. More generally, because inflections like the past tense are simple, frequent, and prevalent across languages, and because the regular and irregular variants can be equated for complexity and meaning, they have served as a test case for issues such as the neurocognitive reality of rules and other symbol-manipulating operations and the interaction between storage and computation in cognitive processing [5–7].

In this article we defend the side of this debate that maintains that rules are indispensable for explaining the past tense, and by extension, language and cognitive processes [3–5,14]. We review what the theory does and doesn’t claim, the relevant evidence, the connectionist challenges, and our hopes for the future of the debate.

The Words-and-Rules theory

The Words and Rules (WR) theory claims that the regular–irregular distinction is an epiphenomenon of the design of the human language faculty, in particular, the distinction between lexicon and grammar made in most traditional theories of language. The lexicon is a subdivision of memory containing (among other things) the thousands of arbitrary sound–meaning pairings that underlie the morphemes and simple words of a language. The grammar is a system of productive, combinatorial operations that assemble morphemes and simple words into complex words, phrases and sentences. Irregular forms are just words, acquired and stored like other words, but with a grammatical feature like ‘past tense’ incorporated into their lexical entries. Regular forms, by contrast, can be productively generated by a rule, just like phrases and sentences. A stored inflected form of a verb blocks the application of the rule to that verb (e.g. brought pre-empts brought). Elsewhere (by default) the rule applies: it concatenates -ed with the symbol ‘v’, and thus can inflect any word categorized as a verb (see Fig. 1).

Irregular forms, then, do not require an ‘exception module’. They arise because the two subsystems overlap in their expressive power: a given combination of features can be expressed by words or rules. Thus either a word (irregular) or a rule-product (regular) can satisfy the demand of a syntactic or semantic representation that a feature such as past tense be overtly expressed. Diachronically, an irregular is born when (for various reasons) learners memorize a complex word outright, rather than parsing it into a stem and an affix that codes the feature autonomously [3].
The WR theory contrasts with classical theories of generative phonology and their descendants, such as those of Chomsky and Halle [15–17], which generate irregular forms by affixing an abstract morpheme to the stem and applying rules that alter the stem’s phonological composition. Such theories are designed to account for the fact that most irregular forms are not completely arbitrary but fall into families displaying patterns, as in ring-rang, sink-sank, sit-sat, and feel-felt, sleep-slept, bleed-bled. A problem for this view is that irregular families admit numerous positive and negative counterexamples and borderline cases, so any set of rules will be complex and laden with exceptions, unless it posits implausibly abstract underlying representations (e.g. rin for run, which allows the verb to undergo the same rules as sing-sang-sung).

The theory also contrasts with the Rumelhart–McClelland model (RMM) and other connectionist models that posit a single pattern associator, with neither lexical entries nor a combinatorial apparatus [1,18,19]. The key to these pattern associators is that rather than linking a word to a word stored in memory, they link sounds to sounds. Because similar words share sounds, their representations are partly superimposed, and any association formed to one is automatically generalized to the others. This allows such models to acquire families of similar forms more easily than arbitrary sets, and to generalize the patterns to new similar words. Having been trained on fling-flung and cling-clung, they may generalize to spling-splung (as children and adults occasionally do [20,21]); and having been trained on flip-flipped and clip-clipped, they generalize to pilp pilled.

WR is descended from a third approach: the lexicalist theories of Jackendoff, Lieber, and others, who recognized that many morphological phenomena are neither arbitrary lists nor fully systematic and productive [22–25]. They posited ‘lexical redundancy rules’, which do not freely generate new forms but merely capture patterns of redundancy in the lexicon, and allow sporadic generalization by analogy. Pinker and Prince proposed that lexical redundancy rules are not rules at all, but consequences of the superpositional nature of memory: similar items are easier to learn than arbitrary sets, and new items resembling old ones tend to inherit their properties. They argued that RMM’s successes came from implementing this feature of memory, and proposed the WR theory as a lexicalist compromise between the generative and connectionist extremes. Irregulars are stored in a lexicon with the superpositional property of pattern associators; regulars can be generated or parsed by rules.

Ullman and colleagues have recently extended the WR theory to a hypothesis about the neurocognitive substrate of lexicon and grammar. According to the Declarative/Procedural (DP) hypothesis [5,26], lexical memory is a subdivision of declarative memory, which stores facts, events and arbitrary relations [27,28]. The consolidation of new declarative memories requires medial-temporal lobe structures, in particular the hippocampus. Long-term retention depends largely on neocortex, especially temporal and temporo-parietal regions; other structures are important for actively retrieving and searching for these memories. Grammatical processing, by contrast, depends on the procedural system, which underlies the learning and control of motor and cognitive skills, particularly those involving sequences [27,28]. It is subserved by the basal ganglia, and by the frontal cortex to which they project – in the case of language, particularly Broca’s area and neighboring anterior cortical regions. Irregular forms must be stored in the lexical portion of declarative memory; regular past-tense forms can be computed in the grammatical portion of the procedural system.

http://lics.trends.com
What the words-and-rules theory does not say

The WR theory does not literally posit the discrete rule ‘to form the past tense, add -ed to the verb’. All it posits is the past-tense morpheme -ed, a variable V’ (included both in the attachment conditions for -ed and the lexical entry of every verb), and a general operation of merging or unifying constituents. The ‘regular rule’ or ‘past-tense rule’ is shorthand for the unification operation applied to the past-tense morpheme. WR is thus compatible with constraint- and construction-based theories of language, as long as they allow for variables and combinatorial operations [29].

WR does not posit that regular forms are never stored, only that they do not have to be [3,30–32]. It would be difficult to prohibit regular forms from ever being stored, given that human memory can acquire many kinds of verbal material (e.g. idioms, clichés, poems). WR posits a parallel-race model, like those defended for inflection by Baayen and Caramazza and by many psycholinguists for visual word recognition [33–39]. Whether a regular form is stored, and whether stored regular forms are accessed, depends on word-, task-, and speaker-specific factors [5,40–43]. For example, regular forms that constitute doublets with irregulars, such as dived/dove and dreamed/dreamt, must be stored to escape blocking by the irregular. As predicted, judgments of the naturalness of regular doublet forms show strong effects of frequency but other regular forms do not [30]. The same is true for regular forms of verbs that resemble irregulars (such as blinked and glided), because the forms must overcome a partial blocking effect exerted by the similar irregulars [30,32]. Tasks that require people to be sensitive to the physical form of words (such as progressive demasking) or to the prior existence of words (such as lexical decision), as opposed to tasks that ask people to judge possible forms, are likely to tap stored representations for medium- and high-frequency regular forms [3,35,44].

Finally, WR is not a chimera of a connectionist pattern associator glued onto a rule system. The lexicon has superpositional properties similar to a pattern associator, but lexical entries have structured semantic, morphological, phonological and syntactic representations of a kind not currently implemented in pattern associators.

Empirical tests

The key predictions of WR are: (1) that irregulars should have the psychological, linguistic and neuropsychological signatures of lexical memory, whereas regulars will often have the signatures of grammatical processing; and (2) that speakers should apply regular inflection whenever memory fails to supply a form for that category. A stored form may be unavailable for many reasons: low or zero frequency, lack of a similar form that could inspire an analogy, inaccessibility because of a word’s exocentric structure (see below), novelty of the form in childhood, and various kinds of damage to the neurological substrate of lexical memory. The heterogeneity of these regular-eliciting circumstances offers converging evidence for distinguishable subsystems, including a productive default that does not critically depend on the statistics of patterns in memory. Here we discuss three types of evidence for a distinction between lookup and concatenation, and connectionists’ attempts to provide alternative accounts (for reviews, see [3,4,14,31]).

Generalization to unusual novel words

The RMM model produced odd blends (mail-membled, triib-treehlit), or no output, for novel words unlike those in its training set [2,20]. People, by contrast, readily apply regular inflections to novel unusual words [20]. According to WR, this is because -ed can attach to any word classified as a verb, even if dissimilar to existing stored regulars.

One connectionist explanation of the difficulties of the model is that they are specific to RMM, which is an early modeling exercise lacking a proper phonological representation, a hidden layer, and a proper output decoder. However, a pattern associator remedying all three deficiencies also had trouble generalizing to unusual words [45]. More recent models that are claimed to solve the problem do so, tellingly, by implementing or presupposing a rule. For example, Hare, Elman and Daugherty installed a ‘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 [46]—in effect, an innate mechanism dedicated to the English past tense. Many recent models have given up on generating past-tense forms; their output layer contains one unit for every past-tense change, turning inflection into a multiple-choice test among a few innate possibilities [47–49].

To convert the choice into an actual form, some other mechanism would have to copy the stem and apply the pattern corresponding to the selected unit. Such a mechanism is simply a rule. Marcus has argued that pattern associators’ difficulty in generalizing to dissimilar forms is rooted in their design [4].

Another response is to claim that people’s success at generalization depends on certain statistical patterns that also foster generalization in pattern associators. Many connectionists claimed that robust generalization depends on regular forms constituting the majority of forms in the child’s input [50]. However, the onset and rate of over-regularization errors in children do not correlate with changes in the number or proportion of regular verbs used by parents [11,51,52]. Moreover, there are regular inflections in other languages, such as the German -s plural, that apply to a minority of nouns (~7%), but are generalized like English regular inflection, namely, to unusual nouns, exocentric nouns, and in childhood [50].

Several modelers now argue that it is not the number or proportion of regular words that is crucial but their distribution in phonological
Box 1. Systematic regularization

An intriguing aspect of inflection is that irregular forms can sometimes turn up in regular form. Some of these regularizations are unsystematic—for example, doublets such as dived/dove and dream/dreamed, in which the regular form is used sporadically because the irregular form is low in frequency and hence poorly remembered. But many are systematic: in particular contexts, the regular form is consistently used, such as ringed the city and low-lifes.

The Words-and-Rules theory explains this phenomenon using an independently motivated theory of compositionality in word-formation [a,b] (see also Fig. 2 in main article). Irregular-sounding words are regularized if they lack a root in head position that can be marked for the inflectional feature (tense or number). The regular suffix applies as the default, as it does in other cases where memory access is disabled. This neatly explains a diverse set of systematic regularizations found in actual usages, laboratory experiments with adults and children, and many languages [c-f]:

The word lacks a noun or a verb root
- onomatopoeia: dinged, pinged, zinged, peeped, beeped
- quotations: ‘I found three man’s on page 1’, ‘We to be’d and not to be’d in this room’
- names: the Julia Childs, the Thomas Manns, the Shelby Footes
- truncations: synched, sysmans
- unassimilated borrowings: tislamians, mongooses

The root cannot be marked for the feature
- verbs with noun or adjective roots: ringed the city, steeled myself, splited the pig, bared his soul, righted the boat, stringed the peas
- nouns with verb roots: a few loafs (episodes of loafing), a couple of wolfs (wolling down food)

The word’s structure is excentric
- verbs based on nouns based on verbs: grandstanced, filled out, costed out the grant, encasted his leg
- nouns based on names based on nouns: Mickey Mouses (simpletons), Renault Elfs, Top Shells (frozen food), Seaowlf (aircraft), Toronto Maple Leafs
- nouns whose referents are distinct from those of their roots: low-lifes, still lifes, sabre-tooths, Walkmans, tenderfoots

Although the meaning of the regularized forms differs from that of their irregular counterparts, regularization is rarely triggered by differences in semantic features alone, as connectionists sometimes suggest [g,h]. If an irregular-sounding word changes in meaning, but retains a root in head position, it stays irregular, no matter how radical the change or opaque the metaphor:

- compositional prefixing: overate, overshott, undid, preshrank, remade, outsold
- non-compositional prefixing: overcame, understood, withdrew, beheld, withstood, undertook
- compounding: boogeymen, superwomen, muskoxen, stepchildren, milkteeth
- metaphors: straw men, chesmen, snowmen, sawteeth, metrical feet, six feet tall, brachiochildren, children of a lesser god, beewolves, wolves in sheep’s clothing
- idioms: went out with (dated), went nuts (demented), went in for (chose), went off (exploded), went off (spoiled), took in (swindled), took off (launched), took in (welcomed), took over (usurped), took up, (commenced), took a leak (urinated), took a bath (lost money), took a bath (bathed), took a walk (walked), blew over (ended), blew away (assassinated), blew away (impressed), blew up (exploded), blew up (inflated), blew off (dismissed), blew in (arrived)

[scores of other examples with come, do, have, get, set, put, stand, throw, etc.]

References
e Marcus, G.F. et al. (1995) German inflection: the exception that proves the rule. Cogn. Psychol. 29, 189–256

http://trends.com
word's meaning changes, the irregular form is in fact *retained*, such as metaphors (*straw men/*mams, sawteeth, God's children*) and idioms (*cut/*cutted a deal, took a leak, hit the fan, put them down*) [2,3,50]. Accordingly, experiments have shown that just changing the meaning of an irregular verb does not cause people to switch to the regular [60,61]. Although all complex and derived words are semantically different from their bases, when semantic similarity and exocentric structure are unconfounded in a regression, exocentric structure accounts for a significant proportion of the variance in choice of inflectional form, and semantic similarity does not [60].

Equally unpromising is the suggestion that people regularize words to avoid ambiguity [63–65]. Many idioms are ambiguous between literal and idiomatic senses, such as *bought the farm* and *threw it up*, or among different idiomatic senses as well, such as *blew away* (impressed, assassinated), but this does not lead people to switch to a regular to disambiguate one of them (*buoy the farm, threw up*). Conversely, *grandstood* and *low-lives* are unambiguous, but people still find them ungrammatical.

One connectionist model added nodes representing the semantic similarity of the verb to the homophonous noun (e.g. *to ring and a ring*) [64]. The network can then be trained to have these nodes turn off irregular patterns and turn on the regular one. But these unusual nodes are not part of the semantic representation of a verb itself; they are an explicit encoding of the verb's relation to the noun that heads it—that is, a crude implementation of morphological structure. In addition, the modelers had to train the network on regular past tenses of denominal verbs homophonous with irregulars. But such homophones are virtually absent from speech addressed to children, who nonetheless tend to regularize exocentric forms [61].

**Neuropsychological dissociations**

According to WR and DP, damage to the neural substrate for lexical memory should cause a greater impairment of irregular forms (and any regular forms that are dependent on memory storage), and a diminution of the tendency to analogize novel irregular-sounding forms according to stored patterns (as in *spling-splung*). In comparison, damage to the substrate for grammatical combination should cause a greater impairment of the use of the rule in regular forms, and of its generalization to novel forms.

Anomia is an impairment in word finding often associated with damage to left temporal/temporoparietal regions (see Fig. 3a). Patients often produce fluent and largely grammatical speech, suggesting that the lexicon is more impaired than grammatical combination [66]. In elicited past-tense production tasks, patients (compared with controls) do worse with irregular than with regular verbs (Fig. 3b), produce regularization errors like *swimmmed* (which occur when no memorized form comes to mind and the rule applies

or *stood* to percolate up from the entries for *ring* or *stand*. With the irregular form sealed in memory, the suffixation rule steps in as the default, yielding *ringed* and *grandstanded*. Many examples, involving diverse constructions from several language families, have been documented from naturalistic sources and experimentally elicited from children and adults [3,50,60,61]. Apparent counterexamples exist, but virtually all can independently be shown to be cases where people do not assign an exocentric structure to the word [3,60].

There have been three connectionist explanations. One is that if a pattern associator had semantic as well as phonological input units, a complex word with an altered meaning would dilute the associations to irregular forms, favoring the competing regular [62,63]. However, in almost every case in which an irregular
as the default), rarely analogize irregular patterns to novel words (e.g. spling-splung), and are relatively unimpaired at generating novel regular forms like planmed [26, 67, 68]. Agrammatism, by contrast, is an impairment in producing fluent grammatical sequences, and is associated with damage to anterior perisylvian regions of the left hemisphere [69, 70]. As predicted, agrammatic patients show the opposite pattern: more trouble inflecting regular than irregular verbs, a lack of errors like swinned, and great difficulty suffixing novel words [26, 67]. Similar effects have been documented in reading aloud, writing to dictation, repeating and judging words (even when controlling for frequency and length) [67], and in a regular/irregular contrast with Japanese-speaking patients [71].

The predicted double dissociation patterns are also seen in a comparison of neurodegenerative diseases. Alzheimer’s disease (AD) is marked by greater degeneration of medial and neocortical temporal lobe structures than of frontal cortex (particularly Broca’s area) and the basal ganglia, and greater impairment of lexical and conceptual knowledge than of motor and cognitive skills, including aspects of grammatical processing [72]. Parkinson’s disease (PD), associated with basal ganglia degeneration, is marked by greater impairment of motor and cognitive skills (including grammatical processing) than use of words and facts [72, 73]. As predicted, AD patients have more trouble inflecting irregular than regular verbs, are relatively unimpaired at suffixing novel words, generate few irregular analogies for novel words, and produce over-regularization errors; PD patients show the contrasting patterns [26, 32]. Moreover, the performance patterns correlate with the severity of the associated processing impairments in the two populations: anomia in AD, and right-side hypokinesia (an index of left-hemisphere basal ganglia degeneration) in PD [26, 32].

Intriguingly, Huntington’s Disease (HD), caused by degeneration of different basal ganglia structures, results in disinhibition of the projected frontal areas, leading to unsuppressible movements [73]. When HD patients inflect verbs, they show a third pattern: producing extra suffixes for regular and novel words like walkeded, plageded and dugged, but not analogous errors on irregulars like dug or kept – suggesting that these errors are instances of unsuppressed regular suffixation [26, 32].

Converging findings come from other methodologies. In normal subjects, both regular and irregular inflected forms can prime their stems. By hypothesis, a regular form is parsed into affix and stem (which primes itself); an irregular form is associated with its stem, somewhat like semantic priming. Patients with left inferior frontal damage do not show regular priming (walked-walk), although they retain irregular priming (found-find) and semantic priming (swan-goose). A patient with temporal-lobe damage showed the opposite pattern [68, 74, 75]. In studies that have recorded event-related potentials (ERPs) to printed words, when a regular suffix is placed on an irregular word (e.g. the German Muskels) or omitted where it is obligatory (e.g. Yesterday I walk), the electrophysiological response is similar to the Left Anterior Negativity (LAN) commonly seen with syntactic violations. When irregular inflection is illicitly applied (e.g. the German Karusellen) or omitted (e.g. ‘Yesterday I dig’), the response is a central negativity similar to the N400 elicited by lexical anomalies, including pronounceable non-words [40, 76–79]. This suggests that the brain processes regular forms like syntactic combinations and irregular forms like words.

Double dissociations are difficult to explain in pattern associators, because except for artificially small networks, ‘lesioning’ the networks hurts irregular forms more than regular ones [80]. A recent interesting model by Joanisse and Seidenberg...
conceded that distinct subsystems have to be lesioned to produce double dissociations [81]. Although they called these modules ‘phonological’ 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 WR predicts. Moreover, the model failed to duplicate the finding that agrammatic patients have more trouble with regular than irregular verbs [26,67]. 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 [81]. This predicts that patient groups should have parallel tendencies to generalize regular and irregular inflection to novel words (planned and splung, respectively), whereas in fact these tendencies dissociate [32,67]. Finally, the model predicts that selective difficulty with irregular forms should depend on semantic deficits. Miozzo reports an amnestic patient who had difficulty accessing word forms but not word meanings; nonetheless, he had trouble with irregulars but not with regulars [82].

At the same time, the post-RMM connectionist models have revealed the problems in trying to explain all linguistic phenomena with a single pattern-associator architecture. Each model has been tailored to account for one phenomenon explained by the WR theory; unlike RMM, few models account for more than one phenomenon or predict new ones. And models repeatedly build in or presuppose surrogates for the linguistic phenomena they claim to eschew, such as lexical items, morphological structure and concatenation operations. We predict that the need for structured representations and combinatorial operations would assert itself even more strongly if modelers included phenomena that are currently ignored in current simulations, such as syntax and its interaction with inflection, the massively productive combinatorial inflection of polysynthetic languages, and the psychological events concealed by providing the models with correct past-tense forms during training (i.e. children’s ability to recognize an input as a past-tense form, retrieve its stem from memory, compute their own form, and compare the two).

As an increasing number of linguistic and neuropsychological phenomena are addressed, especially the complex data from neuroimaging, inadequacies will no doubt be revealed in both kinds of models. Nothing in linguistics prevents theories from appealing to richer conceptions of memory than simple rote storage. Neither does neural network modeling prohibit structured or abstract representations, combinatorial operations, and subsystems for different kinds of computation. The adversarial nature of scientific debate might sometimes have prevented both sides from acknowledging that features of one model may correspond to constructs of the other, described at a different level of analysis. We suspect that allowing a full range of data to tell us which processes are most naturally explained by which kinds of mechanisms, rather than shoe-horning all phenomena into a single mechanism favored by one or another camp, holds the best hope for an eventual resolution of the past-tense debate.

Acknowledgements
We dedicate this paper to David Rumelhart and Jay McClelland, in acknowledgement of the stimulating effect of their model on psycholinguistic and neurolinguistic research. S.P. was supported by NIH grant HD-18381. M.T.U. was supported by a McDonnell-Pew grant in Cognitive Neuroscience, NSF SBR-9055273, NIH MHS189, and Army DAMD-17-93-V-3018/3019/3020 and DAMD-17-99–2-9007.

The future of the past-tense debate
The Rumelhart–McClelland model was deservedly influential, we believe, because it captured a real phenomenon. The persistence of families of irregular verbs with overlapping partial similarities, and people’s use and occasional generalization of these family patterns according to similarity and frequency, can be simply explained by the assumption that human memory is partly superpositional and associative. Theories that try to explain every instance of redundancy among words using the same combinatorial mechanism used for productive syntax and regular morphology require needless complexity and esoteric representations, and fail to capture the many linguistic, psychological and neuropsychological phenomena in which irregular forms behave like words.

References