Is there a past tense rule?

- Early on, children often produce exceptional past tenses correctly (went, took, etc).
- But at some point, they also produce ‘regularizations’ (‘goed’, ‘taked’)
- Also, children (and adults) produce ‘regular’ inflections for novel items when prompted, as in:
  this man is ricking... yesterday he ____.
- This was once taken as suggesting that young children discover ‘the past tense rule’.
- The fact that children learn exceptions was explained by ‘memorization’ or ‘lexical lookup’.

An Alternative to Assuming that Children ‘Acquired’ the Past-Tense Rule

- Rumelhart and McClelland (1986) proposed that rules are not used in forming past tenses, but rather reflect regularities captured in the connections among units in a connectionist system that learns from examples to produce inflected forms.

Overview

- The RM model introduces the connectionist alternative
- Early critiques and responses lead to...
- The Pinker symbolic, dual mechanism account
- Accumulation of arguments and evidence suggests that there is more support for the connectionist approach.
- A new direction builds on the original RM proposal to address the regularity in exceptions.

The RM Model
Training and Testing Procedure

- **Training:**
  - Present WF pattern representing present tense of verb.
  - Compute WF pattern representing past tense of verb using stochastic sigmoid function.
  - Compare computed past-tense pattern to correct past tense pattern.
  - Adjust connections using Perceptron Convergence Procedure (delta-rule)

- **Testing:**
  - Present WF pattern of present tense of verb.
  - Compute WF pattern.
  - Compare to various alternatives on various measures.
  - OR: Generate output using fixed decoding net.

Training Regime

- First ten epochs use 10 most frequent words only
  - feel, have, make, get, give take come, go, look, need
- Remainder of training uses 10 most frequent plus 400 words of ‘middle frequency’
- Each word is presented once per epoch
- An additional 84 lower-frequency words is saved for generalization testing
Recapitulation of U-shaped learning


graph showing percentage of features correct over trials with curves labeled Regular and Irregular.

Responses to t/d and other verbs


table 10:

<table>
<thead>
<tr>
<th>Verb Ending</th>
<th>Regular Suffix</th>
<th>No Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not t/d</td>
<td>203</td>
<td>34</td>
</tr>
<tr>
<td>t/d</td>
<td>42</td>
<td>157</td>
</tr>
</tbody>
</table>

Table 11:

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Verb Ending</th>
<th>Regularized</th>
<th>No Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-15</td>
<td>not t/d</td>
<td>0.44</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>t/d</td>
<td>0.35</td>
<td>0.27</td>
</tr>
<tr>
<td>16-20</td>
<td>not t/d</td>
<td>0.32</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>t/d</td>
<td>0.25</td>
<td>0.35</td>
</tr>
<tr>
<td>21-30</td>
<td>not t/d</td>
<td>0.52</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>t/d</td>
<td>0.32</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Performance on Irregulars by Type

Table 14:

<table>
<thead>
<tr>
<th>Example</th>
<th>Rank Order</th>
<th>Type</th>
<th>Percent</th>
<th>Type</th>
<th>Ratio</th>
<th>Type</th>
<th>Ratio</th>
<th>Type</th>
<th>Ratio</th>
<th>Type</th>
<th>Ratio</th>
<th>Average Trials 11-30</th>
</tr>
</thead>
<tbody>
<tr>
<td>blew</td>
<td>1</td>
<td>VIII</td>
<td>80</td>
<td>VIII</td>
<td>.86</td>
<td>VIII</td>
<td>.76</td>
<td>VIII</td>
<td>.61</td>
<td>VIII</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>sang</td>
<td>2</td>
<td>VI</td>
<td>55</td>
<td>VI</td>
<td>.80</td>
<td>VII</td>
<td>.74</td>
<td>VII</td>
<td>.61</td>
<td>VII</td>
<td>.69</td>
<td></td>
</tr>
<tr>
<td>bit</td>
<td>3</td>
<td>V</td>
<td>34</td>
<td>V</td>
<td>.76</td>
<td>V</td>
<td>.60</td>
<td>IV</td>
<td>.48</td>
<td>V</td>
<td>.56</td>
<td></td>
</tr>
<tr>
<td>broke</td>
<td>4</td>
<td>VII</td>
<td>32</td>
<td>V</td>
<td>.72</td>
<td>IV</td>
<td>.59</td>
<td>V</td>
<td>.46</td>
<td>IV</td>
<td>.56</td>
<td></td>
</tr>
<tr>
<td>felt</td>
<td>5</td>
<td>III</td>
<td>13</td>
<td>IV</td>
<td>.69</td>
<td>III</td>
<td>.57</td>
<td>III</td>
<td>.44</td>
<td>III</td>
<td>.53</td>
<td></td>
</tr>
<tr>
<td>caught</td>
<td>6</td>
<td>IV</td>
<td>10</td>
<td>III</td>
<td>.67</td>
<td>VI</td>
<td>.52</td>
<td>VI</td>
<td>.40</td>
<td>VI</td>
<td>.52</td>
<td></td>
</tr>
</tbody>
</table>
### Performance with Novel Irregulars

<table>
<thead>
<tr>
<th>Verb Type</th>
<th>Presented Word</th>
<th>Phonetic Input</th>
<th>Phonetic Response</th>
<th>English Rendition</th>
<th>Response Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>bid</td>
<td>/bɪd/</td>
<td>/bɪd/</td>
<td>(bid)</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>thrust</td>
<td>/θrɪst/</td>
<td>/θrɪst/</td>
<td>(thrusted)</td>
<td>0.57</td>
</tr>
<tr>
<td>II</td>
<td>bend</td>
<td>/bɛnd/</td>
<td>/bɛnd/</td>
<td>(bended)</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>lend</td>
<td>/lɛnd/</td>
<td>/lɛnd/</td>
<td>(lended)</td>
<td>0.70</td>
</tr>
<tr>
<td>III</td>
<td>creep</td>
<td>/krep/</td>
<td>/krep/</td>
<td>(creeped)</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>weep</td>
<td>/wɛp/</td>
<td>/wɛp/</td>
<td>(wept)</td>
<td>0.34</td>
</tr>
<tr>
<td>IV</td>
<td>catch</td>
<td>/kæt/</td>
<td>/kæt/</td>
<td>(caught)</td>
<td>0.67</td>
</tr>
<tr>
<td>V</td>
<td>breed</td>
<td>/breɪd/</td>
<td>/breɪd/</td>
<td>(bred)</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>grind</td>
<td>/grɪnd/</td>
<td>/grɪnd/</td>
<td>(grind)</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>wind</td>
<td>/wɪnd/</td>
<td>/wɪnd/</td>
<td>(wind)</td>
<td>0.37</td>
</tr>
<tr>
<td>VI</td>
<td>cling</td>
<td>/klaɪn/</td>
<td>/klaɪn/</td>
<td>(clinged)</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>dig</td>
<td>/dɪg/</td>
<td>/dɪg/</td>
<td>(dug)</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>stick</td>
<td>/stɪk/</td>
<td>/stɪk/</td>
<td>(sticked)</td>
<td>0.53</td>
</tr>
<tr>
<td>VII</td>
<td>tear</td>
<td>/tɛr/</td>
<td>/tɛrd/</td>
<td>(teared)</td>
<td>0.90</td>
</tr>
</tbody>
</table>

### Novel Regulars

48/72 only activated correct responses; 6 activated no response; these are the remaining 18 items.

### Summary
- Model can learn regulars and exceptions.
- Correctly inflects most unfamiliar regular verbs.
- Also captures children’s tendency to produce occasional ‘irregularization’ responses and other signs of sensitivity to sub-regularities.
- Produces U-shaped developmental curve.

### Critique (Pinker and Prince, 1988)
- Training regime unrealistic
  - Child’s experience is relatively constant over time.
- Performance on regulars not good enough
  - Makes quite a few errors, some quite strange
- Model can’t produce different past tenses for homophones
  - ring the bell, ring the city, wring the clothes
- Wickelfeature representation has problems
Reply: Implementations are not Conceptualizations (MacWhinney & Leinbach, 1991)

- Included semantic as well as phonological input
- Used a different input representation that led to better performance on regulars
- Did not address U-shaped curve

Pinker (1991, and elsewhere)

- Noted that performance on exceptions does show some signs of exhibiting features like those seen in the RM model.
- Proposed a dual mechanism account in which there is one system that uses rules and another that uses an ‘associative memory mechanism’ much like the RM model.
- With Marcus developed the notion that the rule is completely insensitive to semantic and phonological factors, depending only on the form-class of the stem.
- Has waffled extensively on the question of whether the rule is acquired ‘suddenly’.

Is the onset of the regular past tense sudden?

- According to Marcus et al. (1992), it is sudden:
  
  “Adam’s first over-regularization occurred during a three-month period in which regular marking increased from 0 to 100%”

Plunkett and Marchman (1993, 1996)

- Used simplified corpus and network (all present tense forms reduced to three slots, like ‘run’ or ‘put’)
- Found ‘micro-U’ shaped learning
  - Performance on a given item can vacillate so that correct responses precede incorrect responses.
- Noted special difficulty learning ‘arbitrary suppletions’ like go-went and pointed out that they are also very rare in English and other languages, consistent with the properties of connectionist networks.
Let’s see the rest of the picture

- Hoeffner (1996, Ph.D thesis) notes one could just as easily say: “Adam’s first over-regularization occurred during a 6-month period in which regular marking went from 24% to 44%”.

Other Empirical Claims in Pinker (1991)

- Claimed to demonstrate strong dissociations between regulars and exceptions:
  - Performance on exceptions but not regulars is frequency sensitive.
  - Performance on exceptions but not regulars depends on phonological similarity to known exceptions.
  - Brain damage and developmental disorders can selectively impair performance on regulars and irregulars.

Performance of regulars but not exceptions is frequency sensitive

- Connectionist models show much less frequency sensitivity for regulars than exceptions, as illustrated in SM model of single word reading.
- This arises from the fact that regulars benefit from help from what is learned about other words.
- There is ongoing debate about whether a small effect of frequency actually exists among regulars, once ‘special factors’ have been controlled.
- Thus, the evidence here offers no special support for Pinker’s theory.
Phonological similarity to known regulars

- Prasada and Pinker (1993) compared judgments and generation of inflected forms such as plipped (near known regulars) and ploamphed (far).
  - Ploamphed was judged less acceptable and generation slower than plipped.
  - P&P claimed this was due to an influence from phonological features of the stem; when they subtracted stem acceptability/reading time, no difference remained.

- Albright and Hayes (2003) pointed out that this did not provide unambiguous support for their hypothesis.
  - Found strings that were very high in phonological acceptability but differed in whether they had regular or exceptional neighbors.
  - Number of regular and exception neighbors both made independent contributions to ratings and past tense generation time.

Dissociation in a developmental disorder (the case of the ‘grammar gene’)

- Gopnik & Craigo (1991) reported a selective impairment in regular but not exception inflection in the KE family, a large family with a genetically transmitted speech and language disorder.

- Vargha-Khadem et al. (1998) performed a more detailed investigation of the KE family and found:
  - General deficits including nearly all aspects of verbal and non-verbal abilities.
  - Severe orofacial apraxia.
  - Equivalent deficits in regular and exception past-tense formation.

Semantic but not derivational factors affect choice of regular vs. irregular past tense (Ramscar, 2002)

- Both affected and unaffected members of the KE family were tested using a version of Berko’s sentence completion test, with a set of 20 items provided by K. Patterson.
  - Affected individuals were impaired on both types of items.
  - 41% of the exception errors of affected individuals were regularizations, demonstrating sensitivity to the regular past tense.

KE Family Performance on Regular and Exception Verbs

- Both affected and unaffected members of the KE family were tested using a version of Berko’s sentence completion test, with a set of 20 items provided by K. Patterson.
  - Affected individuals were impaired on both types of items.
  - 41% of the exception errors of affected individuals were regularizations, demonstrating sensitivity to the regular past tense.
What about effects of brain damage?

Ullman et al. (1997) considered effects of anterior vs posterior lesions in the Berko sentence completion task.

The effect of posterior lesions is also observed in patients with semantic dementia.

What about the deficit in regular inflection seen in anterior aphasia?

- Lesions to phonology in the J&S model produce a disadvantage for novel verbs, but do not produce an advantage for exceptions over regulars.
- Bird et al. (2003) argued that the apparent advantage for exceptions reflects phonological differences between regular and exceptional past tenses.

A single-mechanism account

- Joanisse and Seidenberg suggest that computation of inflections involves both semantic and phonological representations.
- A deficit in semantics will influence exceptions and lead to regularization errors because semantics provides a source of differentiating information that helps overcome the tendency of the speech input-output pathway to regularize.
- J&S were able to simulate the effect of semantic lesions (although they used 'localist' semantics).

Phonological Complexity Differences between Regular and Exceptional Past Tenses

- The regular past tense always increases the complexity of the word.
  - like -> liked, love -> loved, hate -> hated
- Some forms so created violate phonotactic constraints on mono-morphemic English word forms (Burzio, 1998).
  - Voiced stop-stop pairs (lobbed) never occur
  - Unvoiced stop-stop pairs (as in liked) never occur after diphthongs (fact, but not *faict)
- Exceptional past tenses are generally no more complex than their stems, which are often very simple.
  - eat -> ate, take -> took
  - weep -> wept reduces stem to compensate for added 't'.
Status of Empirical Claims in Pinker 1991

- Claimed to demonstrate strong dissociations between regulars and exceptions:
  - Performance on exceptions but not regulars is frequency sensitive.
  - Performance on exceptions but not regulars depends on semantic and phonological similarity to known exceptions.
  - Brain damage and developmental disorders can selectively impair performance on regulars and irregulars.
- Also claimed that syntactic but not semantic variables affect choice of regular vs. exception.
  - Denominal status: ‘Why no mere mortal ever flew out to left field’ (to ‘fly’ said to be derived from a noun).

Traditional view of morphology

- Word are built out of discrete units (morphemes) that contribute systematically to grammar/meaning (WALKED = WALK + -ED; TEACHER = TEACH + -ER)
  - But systematicity is only partial (SING ⇒ SANG; CORNER ≠ CORN + -ER)
- Dual-mechanism theories (e.g., Pinker, 1999)
  - Regular/transparent words processed compositionally in terms of stems and affixes (e.g., by explicit rules)
  - Irregular/opaque words stored separately as whole forms (e.g., in an associative system)

Problem

- Systematicity is not only partial but graded—many intermediate cases are neither completely systematic nor completely arbitrary
  - No-change verbs (HIT, SPREAD) all end in /t/ or /d/
  - DRESSER not “someone who dresses” but related to dressing
Connectionist morphology

- Morphology is a characterization of the learned mapping among surface forms of words (phonology, orthography) and their meanings (semantics).
  - Internal representations come to reflect graded systematicity across items.
- Properties of morphology can be derived from the nature of semantics, phonology, orthography, and their interrelationships.
  - Still an important level of analysis, but no need to invoke independent principles.
- Performance on lexical tasks should generally be sensitive to graded effects of both semantic and formal similarity, although:
  - The effects may interact.
  - Degree of sensitivity may be language-specific.

Independent effects of morphology (in English)?

- Priming based on morphological type (Marslen-Wilson et al., 1994)
  - M+S+ GOVERNMENT → GOVERN 41 msec *
  - M+S+ GOVERNMENT → GOVERNOR 11 msec
  - Failed to control for phonological relatedness (Seidenberg & Gonnerman, 2000)
- Bound morphemes (Aronoff, 1976)
  - e.g., MIT PERMIT PERMISSION PERMISSIBLE
  - ADMIT ADMISSION ADMISSIBLE
  - TRANSMIT TRANSMISSION TRANSMISSABLE
  - No significant priming (Feldman & Soltano, 1999; Marslen-Wilson et al., 1994)
- Many studies have failed to control semantic and formal similarity jointly
  (e.g., Fowler et al., 1985; Granger et al., 1991; Marslen-Wilson et al., 1994; Napps, 1989; Napps & Fowler, 1987)

No?

Independent effects of morphology (in Hebrew)?

- Root priming for semantically unrelated items (Bentin & Feldman, 1990; Frost, Forster, & Deutsch, 1997)
  - e.g., KLT M+S+ HAKLATA → TAKLIT 15 msec *
  - (a recording) (a record)
  - M+S− KLITA → TAKLIT 11 msec *
  - (absorption) (a record)
- Effects of “structural” manipulations on word-pattern priming among verbs (Frost, Deutsch, & Forster, 2000)
  - Strong word-pattern priming for words and pseudowords with standard roots
  - No word-pattern priming for words with “weak” (2-consonant) forms of roots
Yes

Simulation 1 (Plaut & Gonnerman, 2000, LCP)

Orthography
- Two-syllable “words” constructed from 100 first syllables and 100 second syllables
  - first syllables are “stems”; 10 of second syllables designated as “affixes”
- Each syllable coded as random binary pattern over 15 units (∼ 1/3 on); no similarity structure among syllables

Semantics
- Each syllable assigned random “meaning” over 50 semantic features
  - First syllable activates 10/50 features; second activates 5/50 features
- “Transparent” meaning of a two-syllable word is superposition (bitwise OR) of meanings of syllables
- Four word classes derived by randomly distorting transparent meanings
  - Transparent 100% of transparent features (RUNNER)
  - Intermediate 67% of transparent features (DRESSER)
  - Distant 33% of transparent features (TENDER)
  - Opaque 0% of transparent features (CORNER)
Two “languages”

Created by generating 12 words from each of 100 stems (1200 words total):

- **Experimental** words (480 from 40 stems)
  - 10 forms derived from affixes (evenly distributed among Transparent, Intermediate, Distant, and Opaque conditions)
  - 2 forms “derived” by combining with random second syllables (with Opaque semantics)

- **Background** words (720 from 60 stems)
  - **Rich language**: All derived forms are Transparent
  - **Impoverished language**: All derived forms are Opaque

- Rich and Impoverished languages differ only in the background words
  - All comparisons involve only experimental words (identical across languages)

Training and testing procedures

- 30 orthographic units
- 50 semantic units
- 300 hidden units
- Trained with back-propagation separately on each of two languages using identical initial random weights and learning parameters
- Testing for priming by cascading activations (τ = 0.01) and replacing prime with target after 1.0 unit of time (100 updates)
- Same primes and targets used in both languages
- Target RT based on stability criterion (mean change < 0.01); lexical decision assumed to be based on semantic activation (see, e.g., Plaut, 1997, LCP)

Results

Difference in RT following unrelated vs. related (same stem) prime

<table>
<thead>
<tr>
<th>Morphological Transparency</th>
<th>Rich Language</th>
<th>Impoverished Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transparent</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Intermediate</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Distant</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>Opaque</td>
<td>0.4</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Analysis of hidden representation

Mean correlation of hidden representations of words with a common stem

<table>
<thead>
<tr>
<th>Morphological Transparency</th>
<th>Rich language</th>
<th>Impoverished language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transparent</td>
<td>0.45</td>
<td>0.5</td>
</tr>
<tr>
<td>Intermediate</td>
<td>0.5</td>
<td>0.55</td>
</tr>
<tr>
<td>Distant</td>
<td>0.6</td>
<td>0.65</td>
</tr>
<tr>
<td>Opaque</td>
<td>0.65</td>
<td>0.7</td>
</tr>
</tbody>
</table>