Children's Overregularization and Its Implications for Cognition

Gary F. Marcus

Department of Psychology

University of Massachusetts, Amherst

October 1996

Appeared as: **Marcus, G. F. (2000). Children's Overregularization and Its Implications for Cognition. In P. Broeder, & J. Murre (eds).** ***Models of Language Acquisition: Inductive and deductive approaches.*** **Oxford: Oxford University Press, pp 154-176.**

Acknowledgments

For helpful comments on an earlier draft I thank S. Pinker, F. Xu, Z. Zvolenszky, and two anonymous reviewers. The research described in this article was supported by an NDSE Graduate Fellowship and University of Massachusetts Faculty Research Grant to Marcus, NIH 18381 and NSF BNS 91-09766 to Steven Pinker, and an ACLS/DAAD collaborative research grant to Harald Clahsen, Steven Pinker, and Gary Marcus. Address correspondence to Marcus at Department of Psychology, University of Massachusetts, Amherst MA 01003, or by e-mail: marcus@psych.umass.edu.

1. Introduction

A central goal for cognitive science is to discover basic computational mechanisms that are recruited throughout cognition (Gallistel, 1994). In this paper I propose two mechanisms which underlie learning and generalization.

In particular, I propose that humans have both a statistical mechanism which extracts, collects, and tabulates statistical contingencies in the environment, and a rule mechanism which suppresses variation between examples within a category, thereby treating all instances of a category as equal members. Of course, neither mechanism is entirely novel; indeed, much of the debate in cognitive psychology concerns such mechanisms. My goal here is to argue that *both* mechanisms are necessary. Section 2 presents detailed arguments that argue for the existence of both mechanisms within a single narrow linguistic domain, inflection. Section 3 proposes that these same two types of mechanisms are used in many other domains of cognition.

2. Models of Inflection

In this section, I will argue that statistical and rule-based computational mechanisms are both required within one simple domain: the English past tense. Most verbs in English form their past tense regularly, by adding *-ed*, e.g. *walk-walked*. About 180 "irregular" verbs form their past tense idiosyncratically (e.g. *break-broke, go-went*).

2.1 The rule and memory model

The thesis of this section is that irregular inflected forms are produced by a mechanism which is sensitive to resemblance to stored exemplars in a statistically sensitive memory, while regular inflected forms are produced by a rule which can operate independently of resemblance to stored examples.

Since irregulars undergo unpredictable changes, nearly everyone agrees that they are stored in memory. One fact supporting the notion that irregulars are stored is that high-frequency irregulars are judged as better-sounding than low-frequency irregulars (Ullman, 1993).

The memory for irregular verbs cannot, however, be a simple look-up table, in which each verb is independent from all the others. Irregular verbs tend to resemble one another, falling into clusters of stems undergoing similar changes, such as *sing-sang*, *ring-rang*, and *drink-drank*. In addition, most past tense forms resemble their associated stem: stem-past pairs like *sing-sang*, *sleep-slept* and *think-thought* are much more common than suppletive forms like *go-went*. Moreover, adults can generalize irregular patterns, even to invented verbs such as *spling*. Such generalizations depend on resemblance to stored forms. People are more likely, for instance, to generalize the *i-a* pattern seen in *sing-sang* to novel words that closely resemble stored forms (*spling*) than to novel words that don't closely resemble stored forms (*nist*) (Prasada & Pinker, 1993). Thus, it seems likely that irregulars are stored in an associative memory mechanism which is sensitive to a word's resemblance to stored forms.[1]

More controversial is whether regulars are produced by a rule. Of course, in one sense, English past tense formation can be trivially described by a rule: as a general statistical tendency, most English verbs are inflected with the morpheme /ed/. In this sense, it is a rule that planets follow elliptical trajectories. But clearly planets do not consult a rule, and the elements of the rule have no causal status on the planet's 'behavior' (Pylyshyn, 1990). The more interesting question is whether learners or speakers consult (unconsciously) a rule, in which they manipulate internally represented symbols.

My colleagues and I have argued that regulars are indeed computed by a symbolic rule, roughly, add *-ed* to a verb to form its past tense. (Marcus 1995a, 1995b; Marcus, Pinker, Ullman, Hollander, Rosen and Xu, 1992; Marcus, Brinkmann, Clahsen, Wiese, and Pinker, 1995 ; Pinker, 1991; Pinker & Prince, 1988; Prasada and Pinker, 1993). What makes 'add -ed' a *rule* is that it can apply to any word marked by the symbol *verb*. Because rules suppress distinctions between members of a class, generalization of the *-ed* rule is independent of resemblance to stored forms.

Moreover, 'add -ed' is a *default*, which is to say that it applies as a "last resort", whenever access to a system of stored irregular roots fails. For instance, when Berko (1958) asked children to inflect novel verbs (e.g., *This is a man who knows how to wug; he just* __), they readily inflected *wug* as *wugged*. More generally, default inflection applies in at least 21 heterogeneous circumstances (Marcus, et al., 1995) including unusual sounding words (*ploamph-ploamphed*), denominal verbs (*The soldiers ringed/*rang the city*), and onomatopoeia (*The annoying car alarm beeped/*bept all through the night)* -- any time that access to irregular forms in the lexicon is impossible.

Overregularizations

The rule-and-memory model -- irregulars are stored in memory, regulars are produced by a default rule when access to an irregular fails -- provides a simple, elegant explanation of children's overregularization errors (e.g., *breaked*).

Memory for irregular past tense forms, like any memory, is fallible. Whenever a child's memory for irregular past tense forms fails (because a given past tense form is novel or relatively unfamilar to the child), she falls back on the default rule, and adds *-ed*, producing an overregularization. Just as an adult speaker might inflect *shend* ('to shame') as the overregularization *shended*, because the past tense form *shent* is so infrequent as to be unretrievable, the child, failing to retrieve *broke* as the past tense of *break*, would produce *breaked*.

This model makes several predictions. First, the rate of overregularization should be relatively low, since the errors should occur only as the result of retrieval failure. Second, the more often we hear something, the easier it is to remember, so if overregularizations are the consequence of memory failure, they should occur more often with low frequency verbs. Third, families of similar-sounding irregulars like *sing-sang*, *ring-rang*, *drink-drank* should reinforce one another, and thereby resist overregularization. Fourth, overregularization should disappear gradually, as memory retrieval improves through increased exposure (or better retrieval strategies). Fifth, overregularizations should occur only once the child has learned to use the regular past tense in contexts which require it (e.g., *I walked yesterday*); before the child acquires the default, retrieval failures should result in unmarked forms (*I sing yesterday*).

Tests of the Rule and Memory model

To test these predictions, my colleagues and I collected past tense forms from CHILDES, a collection of computerized transcripts of children's conversations (MacWhinney & Snow, 1985). Using computer programs (corroborated by hand-analyses), we extracted 11,521 past tense forms of irregular verbs [2], correct (*sang*) and overregularized (*singed*, *sanged*) from the transcripts of 83 children, ranging in age from 1;3 to 6;6, with most data between 2;0 and 5;2 (Marcus, Pinker, Ullman, Hollander, Rosen, and Xu, 1992).

We calculated the rate of overregularization (OR) as follows:

$$OR = \frac{\#\text{Overregularized-tokens}}{\#\text{Overregularized-tokens} + \#\text{correct-past-tense-tokens}}$$

As predicted, children do not go through a period where all irregular verbs are overregularized in all or even averaging over time (of four children with longitudinal data, none overregularized in a single month at a rate of over 50%; the vast majority of the months were at rates of less than 10%). This low rate of overregularization also is not an artifact of averaging over children (no child had a mean overregularization rate of greater than 24%, most had rates of less than 10%) or averaging over verbs (most verbs were overregularized at low rates). The comparative rarity of overregularizations suggests that the errors stem from a performance error rather than a qualitative grammatical reorganization in which overregularizations replace or freely alternate with correct forms.

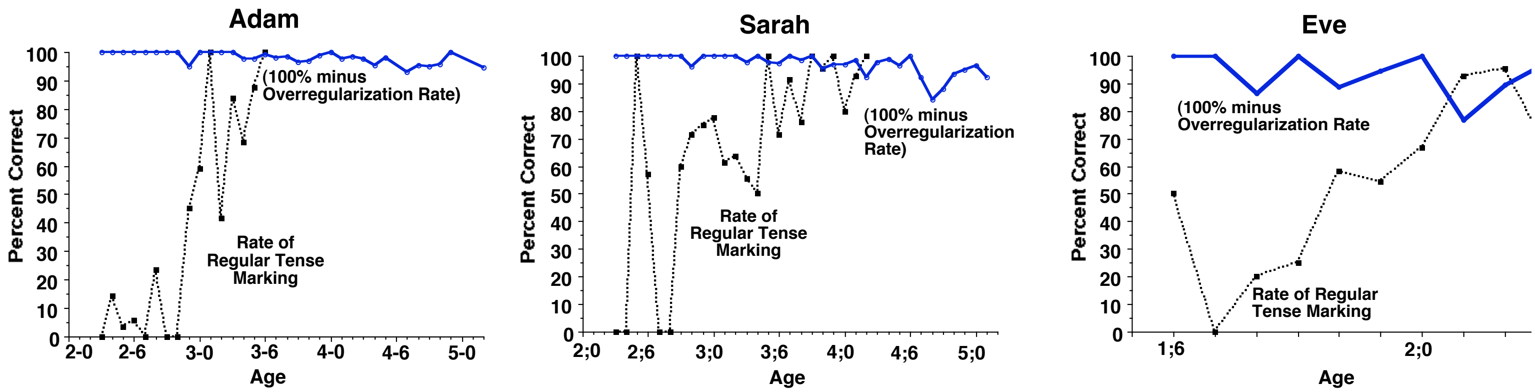

Figure 1: Three children's rate of overregularization subtracted from one, and their corresponding rates of supplying the regular past tense in contexts in which it is required.

To test the second prediction (frequency), we calculated how often verbs were used by parents in the CHILDES transcripts and how often those verbs were used in Francis & Kucera's (1982) standardized frequency database. Both measures confirm Prediction 2: low frequency verbs were more likely to be overregularized than high-frequency verbs (the average correlation between a child's overregularizations rates and his/her parent's frequencies was -.34; 18 of 19 children showed correlations in the same direction), consistent with the notion that overregularizations errors are produced when memory retrieval fails.

Third, we tested the irregular attraction effect by computing a score for each irregular verb based on the number and frequency of similar irregular verbs. We found that the verbs with greater numbers of similar sounding irregular neighbors were less likely to be overregularized. For example, it appears that neighbors such as *drink-drank* and *ring-rang* makes retrieval of *sing-sang* easier, thus rendering *singed* less likely. This suggests that irregular forms are stored in a pattern associator memory in which similar forms reinforce each other. (The converse hypothesis, attraction to regular verbs, will be discussed in the next section.)

The fourth prediction was that if overregularizations disappear through the child's increasing exposure to correct forms, rather than a qualitative grammatical reorganization, they should disappear gradually over time. This appears to be correct: our sample, drawn almost entirely from preschoolers, overregularized at a rate of 4.0%, first-graders 2.5% (Moe, et al., 1982), fourth-graders 1% (Carlton, 1947), and even adults overregularize sometimes, approximately once in every 25,000 opportunities (Stemberger, 1989); this gradual decline is consistent with a gradually increasing memory trace for irregulars.

Finally, before children reliably inflect regular verbs for past tense in contexts that require it (e.g., *I walked yesterday)*, irregular verbs are used correctly or left unmarked, but never overregularized. Instead, consistent with this prediction, Figure 1 (above) shows that the onset of overregularization appears to coincide with the development of reliable regular past tense marking, suggesting that overregularization is triggered by the acquisition of a rule (see

2.2 Single uniform connectionist networks

Despite the success of the rule and memory model, one might imagine a model which would attribute regular inflection to the same storage and retrieval process as irregulars. On such a view, there would be no explicit distinction between "regular" and "irregular"; instead, there would be a continuum between the two. For instance, Bybee (1991:86-87) suggested that "All types of morphological patterns can be acquired by the same process -- the storage of items, the creation of connections among them, and the formation of patterns that range over sets of connections."

This approach has been precisely implemented by single uniform neural (connectionist) networks. These networks make no explicit distinction between regular and irregular words, and contain no explicit rules. Instead words are represented as distributed patterns of activation over a network of nodes and connections. Simplifying a bit, *sing* would be represented as a set of units representing the initial consonant cluster, the vowel nucleus, and the final consonant. The network is presented with pairs of stem and past tense forms. Learning *sing-sang* involves strengthening connections between nodes, such as the connections between nodes representing the sound *ing* and nodes representing the sound *ang*. Similar-sounding words overlap in their representations, yielding an explanation for why adults sometimes inflect the novel word *spling* as *splang* (Bybee & Moder, 1983; Prasada & Pinker, 1993).

Regular inflection is treated identically: exposure to *walk-walked* strengthens the connections between input nodes representing *alk* and output nodes representing *alked*, allowing the network to generalize to similar words like *talk-talked*.

Overregularizations are produced -- without rules -- when regular patterns exert too strong an attraction on irregular verbs. For instance *grow* might be overregularized to *growed*, because words like *glow-glowed* increase the strength of the connection between nodes representing *ow* and *owed*. Other things being equal, as the network is exposed to more such pairs, the chance of overregularization increases.

This statistical dependency is readily demonstrated. For instance, Plunkett and Marchman (1993) experimented with a series of simulations which varied the proportion of regular verbs in the training. They found that "generalization properties [of their network] are best understood by reference to characteristics of the training set", because "the level of generalizations ... [is] closely related to the total number of regular verbs in the vocabulary." This relationship might be dubbed "the regular attraction hypothesis" and it is central to most connectionist models of inflection (Marcus, et al., 1995; see also Hare, et al., 1995; Marcus, submitted a).

The frequency hypothesis can thus be used as a wedge to distinguish pattern associator models from the rule and memory model. Whereas in current connectionist models, regular inflection depends on high regular type frequency; in the rule-based model, default inflection applies to any word carrying the symbol [verb], independent of frequency. The next section presents four tests of this prediction.

Tests of the type frequency hypothesis

Longitudinal test

As noted previously, children's first inflected past tense forms are correct (or unmarked); overregularizations appear only later.

In some connectionist models of inflection, a rapid change in the proportion of regular verbs in the input to the model causes the onset of overregularization. This can be seen most clearly in Rumelhart and McClelland's (1986) model. Their model does show a U-shaped sequence in which overregularizations follow an initial period of correct use, but the onset of overregularization in their model coincides precisely with the shift in training regimes from one in which 80% of the verbs are irregular to one in which 80% of the verbs are regular.

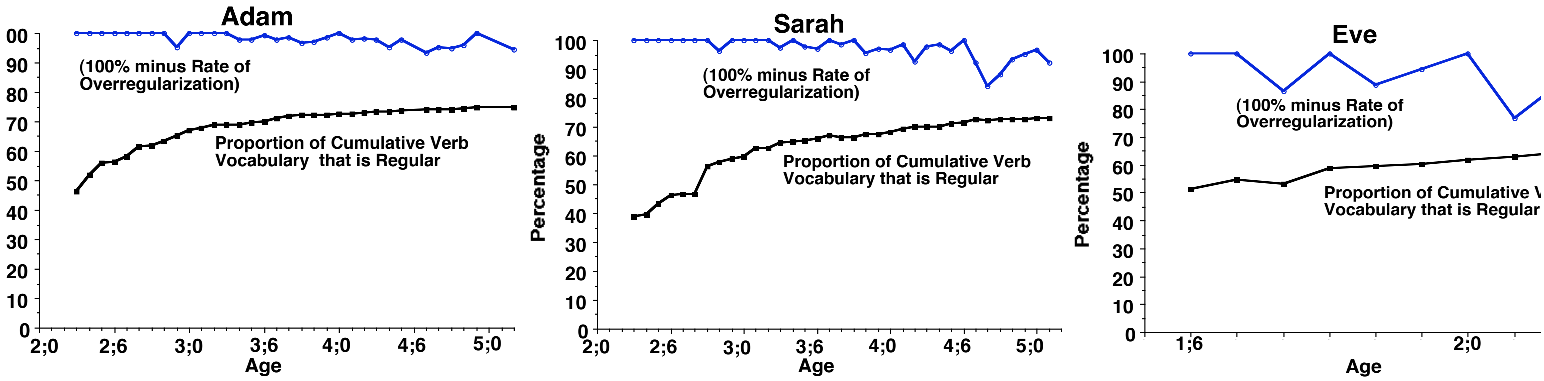

Figure 2: Three children s rate of overregularization (subtracted from 100%) and the proportion of their cumulative verb vocabulary that is regular.

For example, before Adam first overregularizes, all inflected past tense forms of irregular verbs are correct, and simultaneously there is a rapid increase in the proportion of verbs in his vocabulary that are regular. Later, when Adam is overregularizing, the proportion of his verb vocabulary that is regular increases much more slowly, the opposite of the prediction of many connectionist models.

Since this result could conceivably be a statistical artifact of a cumulative vocabulary measure that necessarily increases but decelerates over time, we used “mark-recapture” techniques common in biostatistics to estimate vocabulary as a function of how many verbs types recur from session to session. Roughly, if a child says 50 verb types in sample 1, and 50 verb types in sample 2, if 10% of the verbs overlap, we can estimate that the child’s vocabulary consists of 500 verbs. Estimates constructed this way are free of the bias that could affect the cumulative measures, but yielded similar results (Marcus, et al., 1992, Chapter V). Thus, the timing of overregularization appears to be independent of changes in the proportion of vocabulary that is regular.

Lexical Test

The pattern associator models and the rule and memory model both correctly predict that verbs should be drawn to irregular patterns by similar-sounding irregular verbs. A verb like *sing* will be drawn to *sang* because of *sing*’s overlap with the representational machinery of *ring-rang*, *drink-drank*, and so on. A verb that has many similar neighbors should be easy to remember; a verb that has few neighbors should be hard to remember. As discussed earlier, confirming the predictions of both theories, every child displayed a correlation in the predicted direction: The more irregular neighbors a verb had, the lower its rate of overregularization. This suggests that irregulars are stored in an associative, perhaps PDP-like memory, in which irregulars are not stored as isolated items in a table, but instead exert an attraction on one another.

The two theories differ, however, in their predictions about the attraction to regular verbs. Pattern associator models predict that overregularizations result when an irregular verb is drawn to the regular pattern; *feel* should be drawn to the overregularization *feeled* by regular neighbors such as *heal-healed* or *peel-peeled*. On the other hand, if overregularizations are produced by a rule, rather than through an analogical pull to other regular forms, other regular forms should exert no pull towards overregularization.

We tested the regular attraction effect by computing a score for each irregular verb weighting the number of similar **regular** verbs and their frequencies. Here we found no significant correlations, and no consistent direction; the measures were not related. That is, the presence of similar-sounding regular verbs such as *heal-healed* does not make overregularizations such as *feeled* more likely, consistent with the notion that overregularizations and regular forms are available to children, less than two dozen words take irregular plurals. Since a higher proportion of nouns than verbs are regular, other things being equal, pattern associator models should predict that nouns should be more drawn to overregularization than verbs; the rule and memory model predicts that frequency should have no role, and that rates of overregularization should be unaffected.

To test this, I compared the plural and past tense overregularizations rates in spontaneous speech for 10 children (Marcus, 1995, a). Individual children's past tense overregularization rate (7.5%), were not significantly different from their plural overregularization rate (8.3%), suggesting once again that overregularizations are independent of the proportion of regular types in the language.

<u>Cross-linguistic test</u>

If generalization of regular inflection is truly independent of type frequency, speakers should generalize regular inflection in the same way even when there are far fewer regular words. Consider then the case of the German plural. Pattern associator advocates have sometimes pointed to this system as support for their models. For instance, MacWhinney and Leinbach (1991) argued as follows:

> Consider a system such as the marking of plurality on the German noun. German plurals can be formed using -en, -er, -s, -e or zero as endings, along with possible vowel ablauting. None of these five possible suffixes is statistically predominant (Köpcke, 1988); none can be characterized as being "the regular ending." In a situation such as this, there is simply no regular pattern at all. Are we to draw some sharp line between English and German speakers by claiming that only the former evidence "rule-governed" behavior?

The German plural system is extremely complicated, with 8 combinations of affixes and umlaut. Mugdan (1977) described it with 10 rules and 15 lists of exceptions. As measured by standard word frequency counts, the -s plurals occupy 98% of English noun plural types and 97% of the English noun plural tokens, whereas the -s plural makes up less than 9% of German types and 4% of German tokens (Marcus, et al., 1995).

Nonetheless, the German *-s* plural behaves as a default, applying in virtually the same cases as the English *-s* plural. As Van Dam (1940) put it, the *-s* serves as *Notpluralendung* 'emergency plural ending'. Indeed, the German *-s* plural applies whenever access to irregulars fails, applying to unusual sounding words (*Pleik*), names (*Thomas Manns*/**Thomas Manner*), onomatopoeia (*wau-waus*), and abbreviations (*ABCs*) (Bornschein and Butt, 1987; Janda, 1990; Marcus et al., 1995). Indeed, of 21 circumstances in English which allow default inflection; nearly all behave similarly in German (Marcus, et al., 1995).

Experimentally, both children and adults generalize the German *-s* to novel words with that sound unusual or that are treated as names (Marcus, et al., 1995; Bartke, et al., 1995; Clahsen, et al., 1992). For example, when novel words rhyme with existing irregular words, children (aged 3-6 years) and adults prefer to inflect them irregularly, but when novel words don't rhyme with existing words, analogy becomes weaker, and children and adults prefer to inflect these unusual sounding words with *-s*. Similarly, both children and adults inflect novel words presented as names with *-s*. In sum, such generalizations show that German speakers generalize *-s* independently of its frequency, applying the suffix *-s* by default to words carrying the symbol [noun].

Implications for connectionist models

The above studies argue that regular inflection can be generalized independently of frequency. These results pose difficulties for contemporary single uniform network models of inflection, since in most of these models the tendency towards regular generalization is closely related to the proportion of regular words in the input . (A modular connectionist model that is not a single uniform model, Hare, Elman, and Daugherty, 1995, is discussed below.)

Whether future models can overcome these limitations is unknown. One strategy would vary existing discover just the right values for these parameters. But, in my view, ten years of connectionist modeling has yielded little progress in establishing those values, nor is there yet a systematic theory of how one might find them.

Another possible remedy would depend on discovering a more sophisticated connectionist architecture. For instance, although Rumelhart and McClelland's (1986) model is now widely conceded to be inadequate (for critiques, see Lachter & Bever, 1988; Pinker and Prince, 1988; Bever, 1992; Marcus., et al., 1992; Prasada and Pinker, 1993), many researchers attribute that model's limitations to its now obsolete technology, as in McClelland's (1988) suggestion that "a problem with the [R&M, 1986] past-tense model is that it has no intervening layers of units between the input and the output. This limitation has been overcome by the development of the back-propagation learning algorithm (Rumelhart, et al., 1986). Indeed, the back-propagation algorithm led to a booming cottage industry: advocates of connectionism have proposed at least seven new connectionist models of the past tense, mostly using back-propagation and hidden layers (Cottrell and Plunkett, 1991, Daugherty and Seidenberg, 1992, Gasser and Lee, 1991, Hoeffner, 1992, MacWhinney & Leinbach, 1991, and Plunkett & Marchman 1991, 1993; c.f. Egedi and Sproat, 1991).[3]

But hidden layers and back-propagation have not thus far been a panacea. Back-propagation models still face many of the same difficulties as the original Rumelhart and McClelland network. For instance, none of these models can handle the transition from correct forms to the first overregularizations without resort to implausible external

changes (Marcus, 1995b). As one connectionist advocate (O'Reilly,

1996) recently conceded, "While there have been a number of models since the original work of Rumelhart and McClelland ..., and a considerable refinement of the characterization of the empirical phenomenon and the successes and failures of neural network models, the fact remains that no model has captured the U-shaped curve without reverting to implausible manipulations of the training environment or network training signal."

Likewise, without recourse to implementing symbols, no model has provided a plausible account of generalizing low-frequency defaults to words that do not resemble training words. For example, Goebel and Indefrey (this volume) present a model that does generalize the German -s plural substantially, but it has been tested only on words that resemble existing nouns which require the -s plural, thereby not addressing the central theoretical question: can a network generalize a low frequency plural to words to that do not resemble the input?

In fact, as currently constituted, no "single-mechanism" connectionist model can correctly inflect any word that includes a phonological feature that has not appeared in the set of training examples; a formal proof of this is presented in Marcus (submitted, b). More generally, this proof provides a formal account for why the variations in network parameters thus far explored have not yielded models that can generalize default inflection to words that are genuinely unfamiliar.

Interestingly only one connectionist model, Hare, Elman, and Daugherty (1995) appears to be able to generalize a low-frequency default to unfamalar sounding words, but this model succeeds largely by *implementing* a series of hard-wired (as opposed to learned) connections that are specific to the inflectional system being modeled. The model, shown below in Figure 3, contains two networks, a feedforward network and a "clean-up" network. All of the connections in the clean-up network are innately wired: innate inhibitory links run between the inflected nucleus to the -ed *n*ode and the stem vowel nodes, while innate excitatory links run between *-ed* and the stem. In essence, then, what the clean-up network does is to choose between a possible irregular vowel change and the stem suffixed by *-ed*. If the vowel change produced by the feedforward network is strong, it suppress both -ed and the stem vowel. if the vowel change produced by the feedforward network is weak, it excites the -ed node and the stem vowel node. This model is thus transparent implementation of the rule and memory model, not a refutation of it. (For further discussion, see Marcus, submitted a, b.)

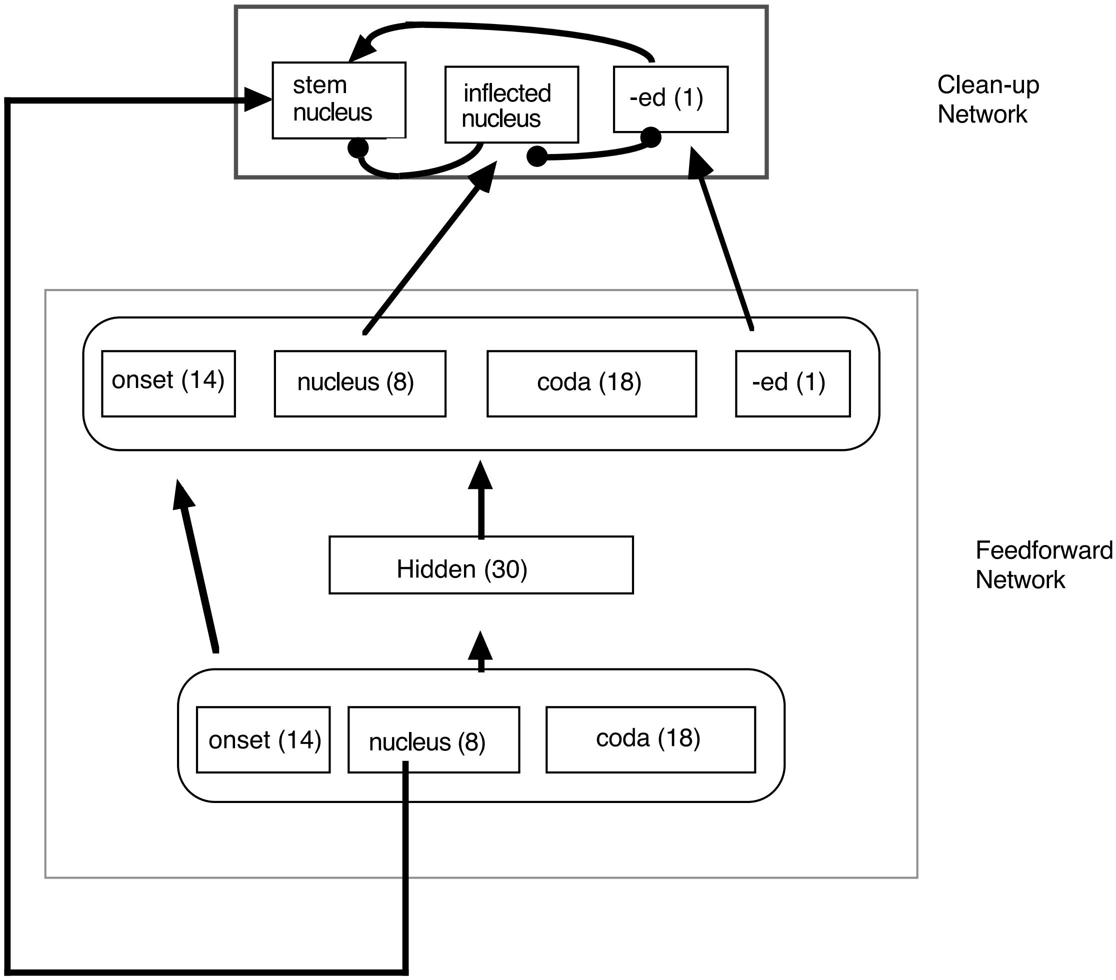

Figure 3: The distributed output representation model presented by Hare et al (1995).

Preliminary results from brain imaging studies support the notion that regulars and irregulars are computed by different mechanisms. For now, no imaging technology is perfect. For instance, psychological studies involving PET (positron emission tomography) have high spatial resolution but low temporal resolution, requiring subjects to repeat items in environments that are not ecologically plausible; in contrast, studies involving ERPs (event-related brain potentials) lack high spatial resolution but have high temporal resolution. Given these limitations it is striking that two recent studies of the production of inflection, one a PET study (Jaeger, Lockwood, Kemmerer, Valin, Murphy, & Khalak, 1996) examining the English past tense, and the other an ERP study (Weyerts, Penke, Dohrn, Clahsen, & Münte, 1996) examining German plurals, yielded essentially similar results in which regulars and irregulars appeared to yield qualitatively different patterns of brain activation. It is difficult to see how single-mechanism models could account for such differences; should these studies prove replicable, they will substantially strengthen the conclusion that regulars and irregular are computed by qualitatively different mechanisms.

2.3 Summary

Comparisons across time, lexical items, inflectional systems, and languages suggest strongly that regular inflection is computed independently of its frequency. We have found that children's onset of overregularizations seems to be tied to the acquisition of a rule, rather than a sudden change in the proportion of regular words in their producing an infinite number of sentences. The lexicon allows us to learn new words; the grammar ensures that even those novel words can be properly inflected and combined.

The fundamental property of a rule is the ability to suppress distinctions between exemplars and treat all exemplars as belonging to a single class. The default rule discussed in this paper, for instance, guarantees that irrelevant information such as frequency or sound can be ignored, focusing on the single key property that the word is a verb. The add *-ed* rule applies just as readily to a frequent, canonical sounding verb like *walk* as to an obscure, infrequent, latinate verb like *perambulate* -- the rule treats *walk* and *perambulate* identically, as words carrying the symbol [verb].

The fundamental property of statistical generalization is sensitivity to quantitative relationships to stored exemplars. Irregular verbs sound better if they are frequent (the irregular change from *sing-sang* sounds better than the corresponding change from *strive-strove*) or if they resemble many other verbs undergoing similar changes (*sing-sang* supports *ring-rang*).

An intriguing possibility is that these two mechanisms serve as basic computational elements recruited throughout cognition, perhaps operating side-by-side within a given domain. While I argue that the two mechanisms are available across domains, I envision that they are more like neural circuits implemented repeatedly throughout the brain than localized subroutines in a central processing unit.

These two mechanisms -- statistics and rules -- are surely not the only computational elements in cognition. For example, I expect that basic arithmetical operations (Gallistel, 1994) and analog imagery processes (Kosslyn, 1994) are also likely to be among the inventory of basic computational elements. Moreover, the nine domains discussed below each use these basic computational elements in different ways, using different rules and tracking the statistics of different entities. Many domains will include domain-specific knowledge and some may also contain mechanisms beyond the two domain-general mechanisms I've proposed here (e.g., Carey, 1985; Cosmides & Tooby, 1994). Still, despite these caveats, I think it may be valuable to rethink these domains. Statistics and rules may both be indispensable devices in the mental tool kit. This section reviews preliminary evidence suggesting that statistical and rule-based mechanisms are used in a broad range of cognitive domains.

3.1 Nine domains in which rules and similarity exist side-by-side

1. Speech perception One of the most compelling demonstrations of people's ability to suppress differences among exemplars and treat all exemplars as stemming from a single class is the phenomenon of categorical speech perception (Liberman, Harris, Hoffman, and Griffith, 1957). People categorize exemplars of /ba/ vs. /pa/ varying in voice onset time into discrete categories. Yet simultaneously, people can discriminate different exemplars of /ba/, and may treat some examples of /ba/ as more representative than other examples (Miller, 1994). Thus both symbolic (categorical) and statistical mechanisms operate within speech perception.

2. Categorization Statistically-based similarity effects in categorization have been well-documented (e.g., Rosch & Mervis, 1975). People can tell that a robin is more representative of the category bird than a penguin, or that a beanbag chair is an unusual chair. Still, evidence accumulated in the last decade shows that people can readily override tremendous perceptual similarity in favor of non-statistical symbolic properties. For example, children can decide that a raccoon surgically transformed to look just like a skunk is still a raccoon (Keil, 1989). Similarly, statistical procedures tell us that 4263 is less prototypical an odd number than 7, but non-statistical procedures still allow us to override this information and treat 4263 as a perfectly good odd number (Armstrong, Gleitman, & Gleitman, 1979). The Pope may be unrepresentative of bachelors; but his bachelorhood does nonetheless guarantee his unmarried status.

3. Gender Constancy The old quip describing androgynous rock stars like Prince, "he sure looks like a girl," amply demonstrates both modes of thinking: the ability to note that a given person, say by the length of his hair, more closely resembles female exemplars coexists with an ability to categorize the person as a male, overriding statistical information (e.g., Bem, 1981).

of a dax to a robin. Still, even inductive reasoning, intrinsically statistical, seems to show a split between symbolic and statistical mechanisms. For example, Carey (1982) found a striking dissociation between similarity and induction. Both children and adults judged a person as being more similar to a mechanical monkey than to a worm, but subjects projected a novel biological property more to the worm than to the toy monkey. These examples demonstrate that induction can override or suppress statistical similarity information in favor of symbolic properties like "is biological." (See Keil, 1989, 1994; Rips, 1989; Smith and Sloman, 1994 for further examples.)

6. Deductive Reasoning. A major area of debate in deductive reasoning asks whether deduction is computed by rule or by analogy to stored exemplars (e.g., Medin & Ross, 1989). Smith, Langston, and Nisbett, (1992) and Sloman (1996) review evidence suggesting that some deductive rules, including modus ponens (if p then q) are treated similarly regardless of the familiarity and concreteness of the content.[4] People are equally comfortable concluding "if p then q. p ?' as 'if it is raining, I'll get wet; it's raining?' More generally the essential property of the syllogism is the ability to suppress differences among exemplars. Consider the (prototypical!) syllogism: *Socrates is a man; all men are mortal*. The very point of this syllogism is that the conclusion "Socrates is mortal" follows no matter how much or how little Socrates resembles other men.

7. *Formal Reasoning* Legal reasoning requires us to ignore superficial resemblance in favor of formal laws (Pinker & Prince, 1989). For example, a bartender may either judge that a mature-looking person looks old enough to drink, based on similarity to stored exemplars, or, perhaps under pressure from a police sting operation, demand a driver's license, focus solely on the patron's age, suppress all other information, and deny even the most responsible, mature-looking 20-year old a drink. Hutchins's (1980) study of the Trobriand islander's legal reasoning suggests that such formal legal reasoning is not restricted to Western cultures, but is instead likely to be a universal property of human nature.

8. *Mathematical reasoning* Children's arithmetic bears a striking resemblance to the past tense. Siegler & Shrager (1984), for instance, have argued that children follow two strategies in solving small addition problems: a quick associative lookup procedure, and if that fails, a slower, rule based counting procedure.

9. *Object recognition* Humans can readily detect similarity to stored exemplars. For example, we readily note changes in illumination, rotation, shadows. Still, one of the most fundamental properties of vision is the ability to recognize an object despite this variability (e.g., Marr, 1982). For example, I can identify my car as my car, under a variety of changes in illumination and orientation. Shepard (1992) argued that one mechanism for suppressing variation between exemplars, color-constancy, has been selected for in order to preserve information about an object's identity under variable lighting and atmospheric conditions. More generally, perceptual constancies, the phenomenological ability to treat exemplars as identical even as retinal size, shape, location, and orientation vary, are among the best known psychological properties -- and depend precisely on the ability to treat an object as a member of a class, regardless of its familiarity or typicality.

3.2 Natural selection, rules, and similarity

Is there any adaptive advantage for an organism developing both of these two mechanisms? Kelly and Martin (1994) argue that sensitivity to statistical information is adaptively advantageous since the environment itself is structured probabilistically. An organism that can detect and exploit probabilistic contingency clearly has an edge over one that can't. Tabulating statistical contingencies also allows an organism to make inferences when no rule applies or when the environment itself is probabilistic.

I concur with Kelly & Martin (1994) that statistical mechanisms are adaptive, providing organisms with a way to cope with a probabilistically structured environment. But I suggest that the rule mechanism is also adaptive: symbols allow for computation in complex algorithms (Dennett, 1995). Symbol-manipulation underlies a calculus of individuation and identification that allows organisms to suppress irrelevant information, focus on relevant information and draw inferences that are likely to be true of every member of the group, even individual members that have not previously been encountered.

The ability to treat an unfamiliar instance as an equal member of a class may underlie three closely-related abilities central to cognition: counting, identification, and individuation (e.g., Starkey, Spelke, & Gelman, 1990; Xu & Carey, 1996). Each demands that we suppress similarity and frequency judgments and treat even an unrepresentative instance as a member. For example, even if a penguin is only 90% as typical a bird as a robin, an organism would be at a serious disadvantage if it were to count five penguins as 4.5 birds!

Judging persistence of identity over time allows us to note that an item retains its identity in a category even as some of its properties change. The abilities to decide that a raccoon surgically altered to look like a skunk is still a raccoon, that a phoneme remains the same even when the pitch or duration varies, that a boy remains a boy even if he wears a hair barrette, that my car remains my car even as the sun goes down and the lighting changes, and that I remain the same person as my hair turns gray -- all depend on constancies that require the ability to suppress statistical information and focus on a symbol -- the hallmark of a rule.

4. Summary

Evidence from a wide variety of cognitive domains suggests that statistical and symbolic mechanisms coexist. It is possible that this commonality is mere coincidence, and that different mechanisms are at work in each domain. Still, the convergence between these domains suggests that two distinct computational mechanisms may be widely available.

The ability to recognize invariance is difficult to capture in purely associative and memory-based systems which tend to be misled by spurious correlations, but well-handled by mechanisms whose operation is opaque to stored exemplars -- precisely the work done by a rule. Correspondingly, rules often fail to capture statistical differences among exemplars -- precisely the work done by statistical generalization. Throughout cognition, the two mechanisms appear to play important, complementary roles.

Psychologists have all too often sought to explain all of mental life with a single mental entity, be it S-R link, rule, or neural network node. But a good theory requires as few mechanisms as needed -- not fewer. As early Scottish psychologist Thomas Reid (1710-1796) noted in *An inquiry into the human mind,* "Men are often led into error by the love of simplicity, which disposes us to reduce things to [a] few principles, and to conceive a greater simplicity in nature than there really is." In my view, cognitive science would be well-served by a modest increase in the mental ontology from one mechanism to two.

References

Armstrong, S., Gleitman, L. & Gleitman, H. (1983) What some concepts might not be. *Cognition,* 13, 263-308.

Bartke, S., Marcus, G. F., Clahsen, H. (1995). Acquiring German Noun Plurals. In D. MacLaughlin & S. McEwan (eds.) *Proceedings of the 19th Annual Boston University Conference on Language Development*. Boston: Cascadilla Press..

Bem, S. L. (1981) Gender schema theory: A cognitive account of sex-typing. *Psychological Review,* 88, 354-364.

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

Bybee, J. L. (1991). Natural morphology: The organization of paradigms and language acquisition. In: T. Huebner and C. Ferguson (Eds.) *Crosscurrents in Second Language Acquisition and Linguistic Theories.* Amsterdam: Benjamins. 67-92.

Carey, S. (1982). Semantic development, state of the art. In E. Wanner and L. R. Gleitman (eds.), *Language Acquisition, State of the Art*. NY: Cambridge University Press.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Carlton, L. E. (1947) Anomalous preterite and past participle forms in the oral language of average fourth grade children. *American Speech*, 22, 40-45.

Cheng, P. and Holyoak, K. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, **17,** 391-416.

Clahsen, H., Rothweiler, M., Woest, A. and Marcus, G. F. (1992). Regular and Irregular Inflection in the Acquisition of German Noun Plurals. *Cognition* **45**, 225-255.

Cosmides, L. and Tooby, J. (1994). Beyond intuition and instinct blindness: toward an evolutionary rigorous cognitive science. *Cognition* **50**, 41-77.

Cottrell, G. W. and Plunkett, K. (1991). Learning the past tense in a recurrent network: Acquiring the mapping from meanings to sounds. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

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

Dennett, D. C (1995). *Darwin's dangerous idea: Evolution and the meanings of life*. NY: Simon & Schuster.

Egedi, D. M., & Sproat, R. W. (1991) Connectionist networks and natural language morphology. Unpublished manuscript, AT&T Bell Laboratories, Linguistics Research Department, Murray Hill, NJ.

Fiske, S. T., Neuberg, S. L. (1990) A continuum of impression formation, from category based to individuating processes: influences of information and motivation on attention and interpretation. In M. P. Zanna (ed). *Advances in Experimental Social Psychology*. San Diego: Academic Press.

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

Gallistel, C. R. Foraging for brain stimulation: toward a neurobiology of computation. *Cognition, 50*, 151-170.

Gasser, M. & Lee, C. D. (1991). A short-term memory architecture for the learning of morphophonemic rules. In R. Lippmann, J. Moody, & D. Touretzky (Eds.), *Advances in neural information processing systems 3*. San Mateo, CA: Morgan Kaufmann.

Halle, M. & Mohanan, K. P. (1985) Segmental phonology of modern English. *Linguistic Inquiry*, 16, 57-116.

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

Hoeffner, J. (1992) Are rules a thing of the past? The acquisition of verbal morphology by an attractor network.

Judd, J. S. (1990). *Neural Network Design and the Complexity of Learning*. Cambridge, MA: MIT Press.

Keil, F. C. (1989) *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: Bradford Books.

Keil, F. C. (1994) Explanation, association, and the acquisition of word meaning. In Gleitman & Landau (eds.), *The acquisition of the lexicon.* Cambridge, MA: MIT Press.

Kelly, M. H. & Martin, S. (1994) Domain-general abilities applied to domain-specific tasks: Sensitivity to probabilities in perception, cognition, and language. In L. R. Gleitman & B. Landau (eds.), *The acquisition of the lexicon.* Cambridge, MA: MIT Press.

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

Köpcke, K.-M. (1988). Schemas in German plural formation. *Lingua 74*: 303-335.

Kosslyn, S. M. (1994). Image and brain : the resolution of the imagery debate. Cambridge, MA: MIT Press.

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.

Liberman, A. M., Harris, K. S. Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, **53**, 358-368.

Ling, C. & Marinov, M. (1993) Answering the connectionist challenge: a symbolic model of learning the past tenses of English verbs. *Cognition, 49,* 235-290.

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

MacWhinney, B. & Snow, C. E. (1985) The Child Language Data Exchange System. *Journal of Child Language*, 12, 271-296.

Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition*, **46**, 53-85.

Marcus, G. F. (1995a). Children's overregularization of English plurals: a quantitative analysis. *Journal of Child Language*, **22**, 447-459.

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

Marcus, G. F. (submitted, a). What does it take to get a connectionist model to generalize a low-frequency default? Manuscript submitted for publication.

Marcus, G. F. (submitted, b). Rethinking eliminative connectionism. Manuscript submitted for publication.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., and Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*. **29**, 186-256.

Marcus, G. F., Pinker, S., & Larkey, L. (1995) The Human Speechome Project: A New Method For Studying Language Development. Poster presented at the Biennial Meeting of the Society for Research in Child Development, March 30-April 2, 1995.

Medin, D. L. & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem solving, and induction. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5). Hillsdale, NJ: Earlbaum.

Miller, J. L. (1994) The internal structure of phonetic categories: A progress report. *Cognition,* 50, 271-285.

Moe, Hopkins, & Rush (1982). The Vocabulary of First-Grade Children. Charles C. Thomas Publisher. Springfield, Ill.

Mugdan, J. (1977). Flexionsmorphologie und Psycholinguistik. Narr: Tübingen.

O'Reilly, R. (1996). *The LEABRA Model of Neural Interactions and Learning in the Neocortex*. Unpublished doctoral dissertation, Carnegie-Mellon University.

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

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

Pinker, S. and Prince, A. (1989) The nature of human concepts: evidence from an unusual source. Unpublished manuscript, Department of Brain and Cognitive Sciences, MIT.

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: acquiring verb morphology in children and connectionist nets. *Cognition*, **48**, 21-69.

Prasada, S. & Pinker, S. (1993) Similarity-based and rule-based generalizations in inflectional morphology. *Language and Cognitive Processses*. **8**, 1-56.

Pylyshyn, Z. (1990). Rules and representations: Chomsky and representational realism. In A. Kasher (Ed.), *The Chomskyian turn*. Cambridge, MA: Blackwell.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou and A. Ortony (eds.) Similarity and analogical reasoning, 21-59. NY: Cambridge University Press.

Rosch, E. & Mervis, C. (1975) Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573-605.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, **323**, 533-536.

Rumelhart, D. and J. McClelland (1986). On learning the past tenses of English verbs. Implicit rules or parallel distributed processing? In: J. McClelland, D. Rumelhart and the PDP research group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. MIT Press: Cambridge, MA.

Shepard, R. (1992). The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In L. J. H. Barkow, L. Cosmides & J. Tooby (eds). *The Adapted Mind*. New York: Oxford University Press.

Siegler, R. S. & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.) *Origins of Cognitive Skills*. Hillsdale, NJ: Earlbaum.

Stemberger, J. P. (1989) The acquisition of morphology: analysis of a symbolic model of language acquisition. Unpublished manuscript, Department of Linguistics, University of Minnesota.

Ullman, M.(1993). *The computation of inflectional morphology*, unpublished doctoral dissertation, MIT.

van Dam, J. (1940) *Handbuch der deutschen sprache*. *Zweiter Band: Wortlehre*. Groningen: J. B. Wolter's uitgevers-Maatschappij N. V.

Wason, P. (1966). Reasoning. In B. Foss (Ed.) *New horizons in psychology*. Harmondsworth, England: Penguin.

Weyerts, H., Penke, M., Dohrn, U., Clahsen, H., & Münte, T. F. (1996). Brain potentials indicate differences between regular and irregular German noun plurals *(Essex research reports in linguistics* No. 13). University of Essex.

Xu, F., & Carey, S. (1996). Infant's metaphysics: The case of numerical identity. *Cognitive Psychology*. **30**, 111-153.

5