Putting these together, we may compute the penalty probability c_{ij} of grammar G_i relative to grammar G_j:

$$c_{ij} = \sum_{G_j \to s} P(s|G_i \not\to s)$$

The pairwise c_{ij}s are given in Table 2.3.

TABLE 2.3. Relative penalty probabilities of the eight grammars

C_{ij}	G_{110}	G_{111}	G_{100}	G_{101}	G_{010}	G_{011}	G_{000}	G_{001}
G_{110}	–	0.790	1.000	0.930	0.750	0.860	0.800	0.930
G_{111}	0.900	–	0.100	0.220	0.750	0.245	0.625	0.395
G_{100}	0.999	0.300	–	0.300	1.000	0.475	0.600	0.475
G_{101}	0.966	0.220	0.100	–	0.750	0.395	0.625	0.245
G_{010}	0.742	0.920	1.000	0.920	–	0.920	0.800	0.920
G_{011}	0.933	0.245	0.325	0.395	0.750	–	0.625	0.220
G_{000}	0.999	0.825	0.750	0.825	1.000	0.825	–	0.825
G_{001}	0.967	0.395	0.325	0.245	0.750	0.200	0.625	–

Currently, we are extending these methods to grammars in a larger parametric space, based on the work of Kohl (1999).

3

Rules over Words

Fuck these irregular verbs.

Quang Phuc Dong, *English Sentences without Overt Grammatical Subject* (1971), p. 4

The acquisition of English past tense has generated much interest and controversy in cognitive science, often pitched as a clash between generative linguistics and connectionism (Rumelhart & McClelland 1986), or even between rationalism and empiricism (Pinker 1999). This is irregular: the problem of past tense, particularly in English, notorious for its impoverished phonology, is a marginal problem in linguistics, and placing it at the center of attention does no justice to the intricacy of the study of language; see e.g. Halle (2000), Yang (2000), and Embick & Marantz (in press).

Yet this is not to say the problem of English past tense is trivial or uninteresting. As we shall see, despite the enthusiasm and efforts on both sides of the debate, there remain many important patterns in the published sources still unknown and unexplained. We show that the variational learning model, instantiated here as competition among phonological rules (rather than grammars/parameters, as in the case of syntactic acquisition), provides a new understanding of how phonology is organized and learned.

3.1 Background

Our problem primarily concerns three systematic patterns in children's acquisition of past tense. First, it has been known since

Berko's (1958) classic work that in general, children (and adults) inflect novel verbs with the *-d* suffix, as in *rick-ricked*. Second, young children sometimes *overregularize*: for example, they produce *take-taked* instead of *take-took*, where the suffix *-d* for regular verbs is used for an irregular verb. On average, overregularization occurs in about 10% of all instances of irregular verbs, according to the most extensive study of past tense acquisition (Marcus et al. 1992). Third, errors such as *bring-brang* and *wipe-wope*, mis-irregularization errors where children misapply and overapply irregular past tense forms, are exceeding rare, accounting for about 0.2% of all instances of irregular verb uses (Xu & Pinker 1995).

One leading approach to the problem of past tense, following the influential work of Rumelhart and McClelland (1986), claims that the systematic patterns noted above emerge from the statistical properties of the input data presented to connectionist networks. A number of problems with the connectionist approach have been identified (e.g. Fodor & Pylyshyn 1988, Lachter & Bever 1988, Pinker & Prince 1988, Marcus et al. 1992). To give just one example (from Prasada & Pinker 1993), connectionist models have difficulty with the *Wug*-test, the hallmark of past tense knowledge. When novel verbs such as *slace* and *smeeb* are presented to a trained connectionist model, *fraced* and *imin* are produced as their respective past tense forms, a behavior hopelessly incompatible with human performance.

In this chapter, we will critically assess another leading approach to the problem of past tense, the Words and Rule (WR) model developed by Pinker and his associates (Pinker 1995, 1999). The WR model claims that the computational system for past tense consists of two components. In the 'rule' component, following the tradition of generative linguistics, regular verbs are inflected by making use of a default phonological rule, which adds *-d* to the root (stem). This explains the productivity of *-d* suffixation to novel verbs. Equally important to the WR model is the Blocking Principle, a traditional idea dating back to Pāṇini. In past tense formation, the Blocking Principle has the effect of forcing the use of a more specific form over a more general form: for example, *sang* is a more specific realization of the past tense of *sing* than *singed*, and is therefore used. Irregular verbs are learned in the 'word' component, which works like a connectionist network, by direct association/memorization of the pairing between a stem and its past tense. The strength of association is conditioned upon the frequencies of irregular verbs that children hear; thus, memorization of irregular verbs takes time and experience to be perfected. When the child's memory for an irregular form fails, the default *-d* form is used. This accounts for the second salient pattern of past tense acquisition: overregularization errors in child language.

Here we will put forward an alternative approach, the Rules and Competition (RC) model. The RC model treats both irregular and regular verbs within a single component of the cognitive system: generative phonology. Like the WR model, we assume the presence of a default rule, which attaches the *-d* suffix to the stem and in principle applies to all verbs. In contrast to the WR model, we claim that irregular past tense is also formed by phonological rules. That is, errors such as overregularization are not memory lapses, but result from failures to apply appropriate *irregular* phonological rules over the default rule.

The RC model derives from the variational approach to language acquisition, which holds that systematic errors in child language are reflections of coexisting hypotheses in competition. These hypotheses are associated with weights, and it is the weights, or the distribution of the grammars, that change during learning from data. For the problem of past tense, the hypothesis space for each irregular verb x includes an irregular rule R, defined over a verb class S of verbs of which x is a member. For example, the rule [*-t* suffixation & Vowel Shortening] applies to irregular verbs such as *lose*, *deal*, and *dream*. The acquisition of x involves a process of competition between R and the default *-d* rule, the latter of which in principle could apply to all verbs, regular and irregular. The child learns from experience that for irregular verbs, irregular rules must apply, and thus the default *-d* rule must not. Before learning is complete, the default rule will be probabilistically accessed, leading to overregularization errors.

Section 3.2 presents the RC model in detail, including a description of the past tense formation rules in the computational system and a learning algorithm that specifies how rules compete. We will also give a learning-theoretic interpretation and revision of the Blocking Principle that underlies the WR model as well as much of generative phonology. Section 3.3 compares the WR and RC models, based on the child acquisition data reported in Marcus et al. (1992). Specifically, we show that children's performance on an irregular verb strongly correlates with the weight of its corresponding phonological rule, which explains a number of class-based patterns in the acquisition of irregular verbs. These patterns receive no explanation under the WR model, to the extent that the WR model is explicitly formulated. Section 3.4 examines, and rejects, the proposal of pairing stem and past tense with analogy or phonological similarity in the WR model, which one might consider a partial remedy for the problems revealed in section 3.3. Section 3.5 gives a critical review of ten arguments in support of the WR model (Pinker 1995). We show that each of them is either empirically flawed or can be accommodated equally well in the RC model.

3.2 A model of rule competition

A central question for a theory of past tense formation, and consequently, for a theory of past tense acquisition, is the following: Should the *-d* rule be considered together with the inflection of the irregular as an integrated computational system, or should they be treated by using different modules of cognition? The approach advocated here is rooted in the first tradition, along the lines pursued in Chomsky & Halle (1968), Halle & Mohanan (1985), and the present-day Distributed Morphology (Halle & Marantz 1993).

> These rules of verbal inflection constitute a continuum of productivity and generality that extends from affixation of the *-ed* suffix in *decide-decided* to total suppletion in *go-went*. . . . In an intermediate class of cases exemplified by verbs like *sing-sang* or *bind-bound* the changes affect only a specific number of verbs. To deal with such cases, the grammar will not contain a plethora of statements such as 'the past tense of *sing* is *sang*, the past tense of *bind* is *bound*,' etc. Rather, it will contain a few rules, each of which determines the stem vowels of a list of verbs specifically marked to undergo the rule in question. (Halle & Mohanan 1985: 104)

This approach differs from the WR model, in which irregular verbs are individually memorized, to the effect of having 'a plethora of statements'.

3.2.1 A simple learning task

Before diving into the details of our model, let's consider a simple learning task, which may help the reader understand the core issues at a conceptual level.

Suppose one is asked to memorize the following sequences of pairs of numbers (x, y):

(37) (2, 4), (3, 4), (4, 8), (5, 10), (6, 7), (7, 8), (8, 16)

Obviously, one strategy to do this is to memorize all the pairs in (37) *by rote*. The learner will store in its memory a list of pairs, as is: (2, 4), (3, 4), etc. However, there is another strategy, which, when available, seems to be a more effective solution. Notice that (37) contains two regularities between the two paired numbers (x, y) that can be formulated as two rules: $y = x + 1$ for $\{3, 6, 7\}$ and $y = 2x$ for $\{2, 4, 5, 8\}$. In the memory of a learner that employs the second strategy, a list of *x*s will be associated with the rule that generates the corresponding *y*s:

(38) a. $\{3, 6, 7\} \mapsto R_{x+1}$
b. $\{2, 4, 5, 8\} \mapsto R_{2x}$

We liken the acquisition of irregular verbs to the number pair learning task described here. The WR model employs the first strategy: irregular verbs are memorized by rote as associated pairs such as *feed-fed, bring-brought, shoot-shot, think-thought*. The RC model, based on a system of generative phonological rules, employs the second strategy such that irregular verbs are organized by rules that apply to a class of individuals:

(39) a. {feed, shoot, . . .} $\mapsto R_{\text{Vowel Shortening}}$
b. {bring, think, . . .} $\mapsto R_{\text{-t suffixation \& Rime} \rightarrow a}$
c. . . .

In an information-theoretic sense, the rule-based strategy, which allows a more 'compact' description of the data, is the more efficient one.[1] The present model is inspired by Morris Halle's idea (e.g. 1983, 1997a) that rules, and abstract representation of phonological structures in general, serve the purpose of saving storage space in the mental lexicon.

Furthermore, there is reason to believe that the rule-based strategy is preferred when verbs (rather than numbers) are involved. While the number-pairing rules can be arbitrary and mentally taxing, the rules for irregular verbs are not. Irregular past tense rules are often well-motivated phonological processes that are abundantly attested in the language. For example, the rule of Vowel Shortening[2] for verbs such *lose, feel,* and *say,* which shortens the long vowel in closed syllables followed by *-d, -ø,* and *-t* suffixes, is attested in many other suffixation processes in English. Therefore, such rules are frequently encountered by and naturally available to the learner.

With this conceptual background, let us move on to the RC model. In what follows, we will describe the properties of the phonological rules for past tense, and how they compete in the process of learning.

3.2.2 *Rules*

The past tense rules in English fall into two broad dimensions: suffixation and readjustment (Chomsky & Halle 1968, Halle 1990). Suffixation attaches one of the three past tense suffixes, *-d, -t,* and *-ø* (null morpheme),[3] to the verb stem. Readjustment rules, mostly vowel-changing processes, further alter the phonological structure of the stem.

We assume, along with the WR model, that as part of innate Universal Grammar, the child language learner is equipped with the knowledge of a default rule, which applies when all else fails. The default rule for English verb past tense is given in (40):

(40) *The default -d rule:*

$$x \xrightarrow{-d} x + \text{-d}$$

Irregular verbs fall into a number of *classes* as they undergo identical or similar suffixation and readjustment processes. Thus, verbs in a class are organized by a shared rule/process. Such a rule is schematically shown in (41), while the rule system for the most common irregular verbs is given in Appendix B.

(41) *Rule R for verb class S:*

$$x \xrightarrow{R} y \text{ where } x \in S = \{x_1, x_2, x_3, \ldots\}$$

For example, the verb class consisting of *lose, deal, feel, keep, sleep,* etc. employs R = [*-t* Suffixation and Vowel Shortening] to form past tense. Suffixation and readjustment rules are generally independent of each other, and are in fact acquired separately. For example, the suffixes in derivational morphology such as *ity, -al,* and *-tion* must be acquired separately, but they all interact with Vowel Shortening, a readjustment rule that applies to closed syllables under many kinds of suffixation, as shown by Myers (1987):

(42) Vowel Shortening in Suffixation
a. [ay]–[ɪ]: divine-divinity
b. [i]–[ɛ]: deep-depth
c. [e]–[æ]: nation-national
d. [o]–[a]: cone-conic
e. [u]–[ʌ]: deduce-deduction

[1] While the saving achieved by the use of rules may not be significant for English irregular verbs—there are only some 150 in all—it becomes dramatic when we move to other languages. This, along with the issue of irregular phonology in other languages, will be discussed in section 3.4.

[2] The term 'Vowel Shortening' is perhaps a a misnomer. The change in the quality of the vowel actually involves shortening as well as lowering. While keeping this technical issue in mind, we will nevertheless continue to call such processes Vowel Shortening; see Myers (1987) and Halle (1998).

[3] See Halle & Marantz (1993) for arguments that the -ø(null) morpheme is 'real'.

It is natural that pervasive rules like Vowel Shortening can be readily built in to the speaker's phonology, and can be used to form verb classes.[4]

Now the conceptual similarities and differences between the WR model and the RC model ought to be clear. It is not the case that the role of memory is completely dispensed with in the RC model. Every theory must have some memory component for irregular verbs: irregularity, by definition, is unpredictable and hence must be memorized, somehow. The difference lies in *how* they are memorized. In the WR model, irregular verbs and their past tense forms are stored as simple associated pairs, and learning is a matter of strengthening their connections. In the RC model, irregular verbs and their past tense forms are related by phonological rules (suffixation and readjustment), as schematically shown in Fig. 3.1.

Once a rule system such as (41) is situated in a model of learning, a number of important questions immediately arise:

(43)
a. Where do rules such as suffixation and readjustment come from?
b. How does the learner determine the default rule (*-d*)?
c. How does the learner know which class a verb belongs to?
d. How do the rules apply to generate past tense verbs?

We postpone (43c) and (43d) until section 3.2.3, while (43a) and (43b) can be addressed together.

For our purposes, we will simply assume that the relevant rules for past tense formation, both the default and the irregular, are available to the child from very early on.[5] That is, the child is able to extract *-t*, *-ø*, and *-d* suffixes from past tense verbs, and can arrive at the appropriate sound-changing readjustment rules that

4 Note that some irregular verbs are conventionally grouped into vowel shifting classes, e.g. ablaut and umlaut, that are not as homogeneous as the Vowel Shortening class. Ablaut and umlaut only designate the *direction* of vowel shifting, e.g. front → back, but leave other articulatory positions, e.g. [± high/low], unspecified. Hence, further refinement is required within these heterogeneous classes (see Appendix B). We will return to the issue of class homogeneity in section 3.4.

5 In the WR model, it is assumed that the default *-d* rule is not available until a little before the child's third birthday (Pinker, 1995). In section 3.5.3, we show that there is little empirical evidence for this view.

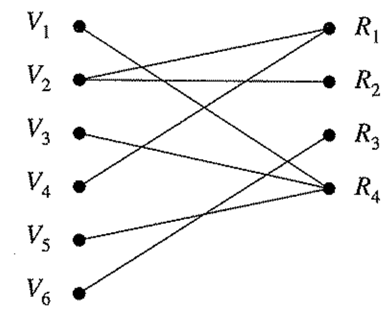

FIGURE 3.1. Verb and rule associations

relate the stem to the derived past tense form. The justification of our assumption is threefold.

First, our assumption is perfectly consistent with children's performance on the past tense. Recall that their past tense is very good (90% correct), and all their errors result from using a wrong rule: almost always the default, very rarely a wrong irregular rule. They do not produce random errors. This suggests that *knowledge* of the rules must be present. What remains problematic, as we shall show later on, is the *application* of these rules.

Second, there is strong crosslinguistic evidence that children's inflectional morphology is in general close to perfect; see Phillips (1995) for a review. For example, Guasti (1992) found that three young Italian children use agreement morphology correctly in more than 95% of all contexts. Clahsen & Penke (1992) had similar findings in a German child during the period of 1;7 to 2;8: the correct use of the affixes *-st* (2nd singular) and *-t* (3rd singular) is consistently above 90%. See Levy & Vainikka (1999) for comparable findings in Hebrew acquisition. And interestingly, when children's morphology occasionally deviates from adult forms, the errors are overwhelmingly of omission, i.e. the use of a default form, rather than substitution, i.e. the use of an incorrect form. This pattern is strikingly similar to that of English past tense learning, where overregularization is far more common than mis-irregularization (Xu & Pinker 1995). To acquire the inflectional morphologies in these languages, the learner must be able to extract the suffixes that correspond to the relevant syntactic/semantic features, and master the readjustment rules and processes when combining stems and suffixes. The learning procedure used there ought to carry over to English past tense.

Finally, recent work in computational modeling of phonological acquisition proposed by Yip & Sussman (1996, 1997) and extended by Molnar (2001) suggests not only that these rules can be learned very rapidly under psychologically plausible assumptions but that they are learnable by precisely the principle of storage minimization. Their system not only learns the correct past tense rules (regular and irregulars), but also learns the correct pluralization rules, at the same time. It learns with far greater efficiency and accuracy than every computational model proposed to date, including MacWhinney & Leinbach (1991), Ling & Marinov (1993), and Mooney & Califf (1995). Since this work is rather technical, we refer the reader to their original papers as well as expositions in Halle & Yang (2002) and Yang (2002).

The rapid learning of rules in the Yip–Sussman model is consistent with the observation that children's knowledge of inflectional morphology is virtually perfect. In section 3.2.3, we lay out the RC model that explains what remains problematic over an extended period of time: the application of these rules.

3.2.3 Rule competition

Class membership

We now return to question (43c), how children learn the class membership of irregular verbs. First, we assume, uncontroversially, that children are able to pair a root with its past tense: for example, when *sat* is heard, the learner is able to deduce from the meaning of the sentence that *sat* is the past tense realization of the root *sit*.[6] Once the root is extracted, the learner can proceed to associate it with the appropriate rule-based class.

[6] For a review that very young children can perform morphological analysis of word structures, see Clark (1993).

It is logically possible that children may put a verb into a wrong class. However, empirical evidence strongly speaks against this possibility. Again, the majority of past tense errors are overregularization errors, which on average occur in about 10% of all instances of irregular verbs (Marcus et al. 1992). Misapplication of irregular rules such as *bring-brang*, *trick-truck*, *wipe-wope*, dubbed 'weird past tense forms' by Xu & Pinker (1995), are exceedingly rare: about 0.2% (ibid.).[7] The rarity of weird past tense forms suggests that the child is conservative in learning verb class membership: without seeing evidence that a verb is irregular, the child generally assumes that it is regular, instead of postulating class membership arbitrarily.

Some notations before we proceed. Write $P(x \in S)$ for the probability that the learner correctly places x into the verb class S. Also, write f_x for the frequency of x in past tense form in the input, and $f_S = \sum_{x \in S} f_x$ for the frequency of a verb *class*, which in the sum of the frequencies of all its members. These frequencies can be estimated from adult-to-child corpora such as CHILDES.

Learning by competition

We now turn to the central component of the RC model: how rules apply to generate past tense verbs, and consequently, how they model the learning behaviors in children's use of irregular verbs.

A central feature of the RC model is that rule application is not absolute. That is, every irregular rule R, which applies to the verb class S, is associated with a weight (or probability) P_S. For example, when the child tries to inflect *sing*, the irregular rule [-ø & ablaut], which would produce *sang*, may apply with a probability that might be less than 1. This follows if learning is gradual: it does not alter its grammar too radically upon the presentation of a single piece of linguistic evidence. If R is probabilistically bypassed, the *-d* rule applies as the default.[8]

[7] See Clahsen & Rothweiler (1993) for similar findings in German acquisition, and Saye & Clahsen (2002) for data in Italian acquisition.

[8] The present model should not be confused with a suggestion in Pinker & Prince (1988), which has an altogether different conception of 'competition'. Pinker & Prince

Now it should be obvious that we have departed from the Blocking Principle assumed in the WR model (Pinker 1995), also known the Elsewhere Condition (Kiparsky 1973) or the Subset Principle (Halle 1997b). The Blocking Principle states that when two rules or lexical items are available to realize a certain set of morphophonological features, the more specific one wins out. For example, *sang* is used to realize the past tense of *sing*, instead of *singed*, because the former is more specific than the latter default rule. Call this version of the Blocking Principle the Absolute Blocking Principle (ABP). In the present model, we suggest a stochastic version of the Blocking Principle (SBP): a more specific rule applies over the default rule *with a probability* (its weight). Thus, a more specific rule can be skipped in favor of a more general rule. The blocking effect of *sang* over *singed* in adult grammar indicates that the weight of the corresponding rule is 1 or very close to 1, *as a result of learning*. In section 3.2.4 we shall return to the Blocking Principle and give empirical arguments for our stochastic version.

An irregular rule *R*, defined over the verb class *S*, applies with probability $P_{\mathcal{R}}$, once a member of *S* is encountered. Thus, it competes with the default *-d* rule, which could apply to an irregular verb, and in fact does, when *R* does not apply. The acquisition of irregular verb past tense proceeds as algorithm shown in Fig. 3.2.

Since regular verbs are almost never irregularized, i.e. the default *-d* rule is almost always employed, let us focus our attention on the case where the verb the learner encounters is an irregular one. When presented with a verb in past tense (X_{past}), the

suggest, much like the present model, that irregular verbs are dealt with by irregular rules (altogether this is not the position they eventually adopt). For them, the competition is *among* the irregular rules the learner postulates: e.g. rules R_1 and R_4 (the target rule) in Fig. 3.1 may compete to apply to the verb V_1. In the present model, the competition is between an irregular and *the default rule*. Under Pinker & Prince's suggestion, when the appropriate irregular rule loses out, another irregular rule will apply. This will result in the very rare mis-irregularization errors: the far more abundant overregularization errors, the main fact in the past tense problem, are not explained.

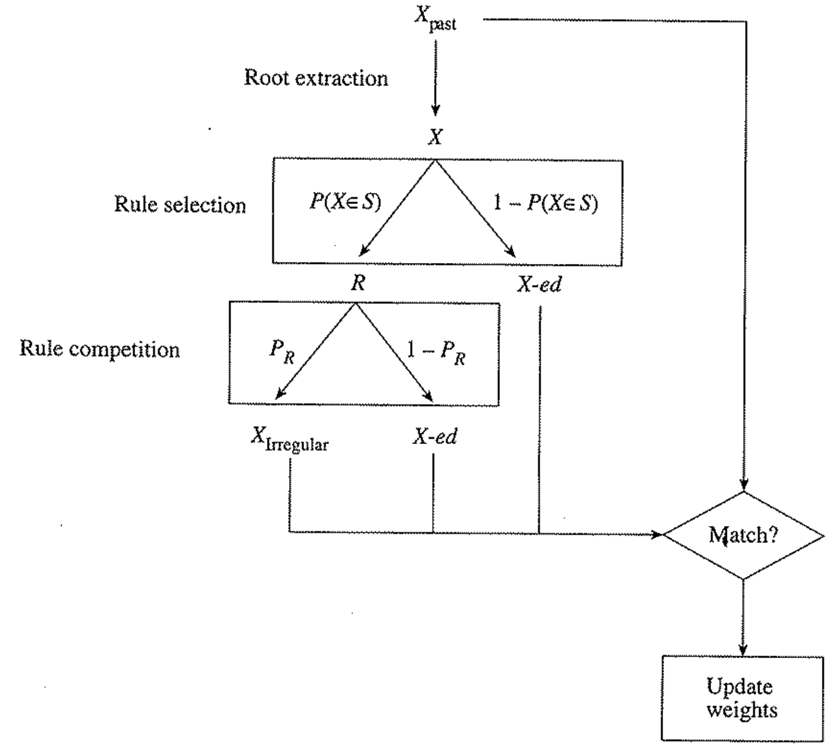

FIGURE 3.2. Learning irregular verbs by rule competition

learner first reconstructs the root *x*. As illustrated in Fig. 3.2, the learner then proceeds to analyze the derivation from *x* to X_{past} in a two-step process:

(44) a. Selection: associate *x* to the corresponding class *S* and hence the rule *R* defined over this class.
b. Competition: apply to *R* to *x* over the default rule.

During learning, either of the two steps may be error-prone. First, the learner may not reliably associate *x* to *S*, in which case *x* would be treated as a regular verb (recall that it is virtually impossible for an irregular verb to be misclassified). That is, in (44a) the probability measure $P(x \in S)$ denotes the likelihood that the learner associates *x* with *S*. Second, even if *x*'s class membership *S* is correctly established, the corresponding rule *R* may not apply:

rather, in (44b), R applies with the probability $P_{\mathcal{R}}$, its weight. Only when both decisions are taken correctly will the correct past tense be produced—a match with the input X_{past}. When either of the two steps fails, the overregularized form will be produced, resulting in a mismatch with the input form, X_{past}.

Thus, for each verb, learning involves updating the two probabilities $P(x \in \mathcal{S})$ and $P_{\mathcal{R}}$. Learning is successful when $\forall x$, $P(x \in \mathcal{S})P_{\mathcal{R}} = 1$: the learner can reliably associate an irregular verb with its matching irregular rule, and reliably apply the rule over the default *-d* rule. As remarked in section 2.3, many models for updating probabilities (weights) are in principle applicable. For our purpose, let us assume a learner who increases the probabilities of the decisions he has made when they lead to a match between the input form and the analyzed form.

Under the null hypothesis, we assume that the grammar system the child uses for production is the same one he uses for comprehension/learning, the two-step procedures in (44). As a result, overregularization of an irregular verb x occurs when either $P(x \in \mathcal{S}) < 1$ or $P_{\mathcal{R}} < 1$.

The RC model makes direct and quantitative predictions about the performance of both irregular verbs and irregular verb classes. Write $C(x)$ to denote the correct usage rate of an irregular verb x; clearly $C(x) = P(x \in \mathcal{S})P_{\mathcal{R}}$. While $P(x \in \mathcal{S})$ may increase when the past tense of x is encountered, $P_{\mathcal{R}}$ may increases whenever *any* member of $\mathcal{S}$ is encountered. These two probabilities, and hence the correct usage of an irregular verb x, is positively correlated with $f_x \times f_{\mathcal{S}}$. Hence, if we hold f_x or $f_{\mathcal{S}}$ constant, the RC model makes two directions about the performance of irregular verbs:

(45) a. For two verbs x_1 and x_2 *within* a verb class, $C(x_1) > C(x_2)$ if $f_{x1} > f_{x2}$.
b. For two verbs x_1 and x_2 such that $x_1 \in \mathcal{S}_1$, $x_2 \in \mathcal{S}_2$, and $f_{x1} = f_{x2}$, $C(x_1) > C(x_2)$ if $f_{\mathcal{S}1} > f_{\mathcal{S}2}$.

In section 3.3 we will systematically evaluate these predictions with children's production data, and demonstrate that irregular verbs are indeed organized into classes.

3.2.4 The Absolute and Stochastic Blocking Principles

We now give justifications for the Stochastic Blocking Principle (SBP), fundamental to the RC model.

Recall that in the WR model, the blocking effect of *sang* over *singed* is given by the ABP: *sang* is used because it is a more specific realization of *sing*+past. The ABP is central to the WR model: when it is presupposed, the rote memorization of irregular verbs is virtually forced. The fact is that children do overregularize, which should be impossible under the ABP. The WR model accounts for this by claiming that that irregular verbs are individually memorized. Overregularization errors are explained by appealing to a principle of association: more exposure leads to better memory. The memory imprints of irregular verbs in a child's mind are not as strong as those in an adult's mind, for children have not seen irregular verbs as many times as adults. Children overregularize because their memory retrieval has not yet become reliable.

Pinker (1995: 112) justifies the ABP by arguing that it is part of the innate endowment of linguistic knowledge, for it cannot be deduced from its effect. His reasoning is as follows. First, to learn the ABP, the child must somehow know that forms like *singed* are ungrammatical. Second, it cannot be concluded that *singed* is ungrammatical from its absence in adult speech—absence of evidence does not imply evidence for absence. Finally, Pinker claims that to know *singed* is ungrammatical 'is to use it and to be corrected, or to get some other negative feedback signals from adults like disapproval, a puzzled look, or a non sequitur response'. Since it is well established (e.g. Brown & Hanlon 1970, Wexler & Culicover 1980, Marcus 1993) that children do not have effective negative evidence, it is concluded that the ABP cannot be learned.

It is not the logic of this argument that we are not challenging; rather, it is the premise that the blocking effect of a more specific form over a more general form is absolute. We show that the effect of the blocking in adult language, the motivation for the Blocking Principle in the first place, can be duplicated as a result of learning, without negative evidence, under our stochastic version of the Blocking Principle.

Suppose that, initially, for the verb *sing*, the irregular rule R=[-ø & ablaut] and the default *-d* rule are undifferentiated. Upon presentation of the past tense form *sang*, both rules have a positive probability of being selected to realize *sing*+past. However, only when R is selected can a match result, which in turn increases its weight (probability), $P_{\mathcal{R}}$. In the end, $P_{\mathcal{R}}$ becomes 1, so that *singed* will never be produced. The end product of such a competition process is a rule system that *appears* to obey the ABP but does not presuppose it: while the specific rule has priority—just as in the ABP—this preference is probabilistic, and gradually increases as a result of learning from experience. In the adult system, the default rule simply does not get a chance to apply, for the more specific irregular rule applies first, and with probability 1.

If the effect of the ABP can be duplicated by rule competition and statistical learning, its theoretical status needs to be reconsidered. Our second objection to the ABP is an empirical one. There is at least one good reason to reject the ABP: the presence of 'doublets'. For example, *learn*+past can be realized as either *learned* or *learnt*, *dive*+past can be realized as either *dived* or *dove*. For doublets, the ABP cannot be literally true, for otherwise *learned* and *dived* should never be possible, blocked by the more specific *learnt* and *dove*. However, the doublet phenomenon straightforwardly falls out of the SBP with a minor change to the learning algorithm: we suppose that the learner punishes P_x when an expected irregular verb x turns out to have regular forms. The term 'expected' is important here, implying that the learner has indeed seen irregular forms of x before, but is now being confronted with conflicting evidence. Presumably, speakers that allow both *learned* and *learnt* encounter and use both forms.[9] As

[9] Including no less a literary genius than Lewis Carroll. In *Alice's Adventures in Wonderland*, *learnt* and *learned* appear exactly once each:

'Yes,' said Alice, 'we learned French and music.'

'Well, I can't show it you myself,' the Mock Turtle said: 'I'm too stiff. And the Gryphon never learnt it.'

a result of competition, the membership probability of *learn* in the corresponding irregular verb class will settle somewhere between 0 and 1, making alternating forms possible.

3.3 Words vs. rules in overregularization

In this section we examine children's overregularization data in detail. We show that the acquisition of irregular verbs shows strong class-based patterns, as predicted by the RC model and the rule-based approach to past tense in generative phonology.

3.3.1 The mechanics of the WR model

In order to contrast the RC model with the WR model, we must be explicit about how the WR model works and what predictions it makes. In the RC model, for any two irregular verbs, we have a concrete claim about their performance in children, based on their input frequencies and the collective frequencies of their respective classes (45), both of which can be estimated from corpora. It is not clear how predictions can be made with this level of clarity under the WR model. Since irregular verbs are learned by associative pairing in the WR model, it is crucial to have a precise statement of how such associative pairing is established. However, the closest to a clear statement that we can find in the WR literature is still vague:

> It is not clear exactly what kind of associative memory fosters just the kinds of analogies that speakers are fond of. Possibly a network of word-word associations might give rise to the right generalization structure if the design of the lexical representation is informed by modern linguistic theory and its implementation is informed by models of superpositional memory. Here we can only present a rough sketch.
>
> Words might be represented in a hierarchical hardware representation that separates stems and affixes, and furthermore distinguishes foot- and syllable-internal structure, finally representing segmental and featural composition at the lowest level of units. Furthermore each of the possible contents of each representation would be implemented once as a single hardware 'type'; particular words would be representation in separate 'token' units with pointers to the types it contains. Links between stems and pasts would be set up during learning between their representations at two levels:

between the token representations of each pair member, and their type representations at the level of representation that is ordinarily accessed by morphology: syllables, onsets, rhymes, feet (specifically, the structures manipulated in reduplicative and templatic systems, as shown in the ongoing work of McCarthy and Prince and others). Ordinary correct retrieval results from successful traversal of token-token links; this would exhaust the process for pairs like *go-went* but would be reinforced by type-type links for members of consistent and high-frequencies families like *sing-sang*. On occasions where token-token links are noisy or inaccessible and retrieval fails, the type-type links would yield an output that has some probability of being correct, and some probability of being an analogical extension (e.g., *brang*). Because the representation of input and output are each highly structured, such extensions would nonetheless be precise and follow constrained patterns, e.g., preserving portions of the stem such as onsets while substituting the appropriate rhymes, and avoiding the chimeras and fuzzy approximations that we do not see among real irregulars but that pure feature-to-feature networks are prone to making. (Pinker & Prince 1994: 334)

It is difficult to evaluate statements like these. The token level association is clear enough: the strength of brute force linking between a stem and its past, hence the retrieval rate of the corresponding verb, can be measured by estimating the frequency of the verb's occurrences in past tense. However, it is not clear how the type-level linkings between phonological structures (syllables, onsets, etc.) are established. But far worse is the vagueness concerning how the two levels interact. For example, while the token-level frequency effect is an important factor in past tense acquisition,[10] it is not clear when the type-level analogy becomes the operative force. Such imprecise formulations are not amenable to analytical results such as (45).

However, we believe that the evidence presented here is strong enough to rule out *any* model that does not use (irregular) phonological rules to describe irregular verbs. The data clearly point to an organization of irregular verbs by rules and classes.

[10] In fact, all 10 pieces of evidence offered by Pinker (1995) in support of the WR model, which we shall review in section 3.5, are frequency based, although section 3.3 has shown that frequency affects performance in a fairly subtle way, unexpected in the WR model.

3.3.2 The data

The measure of children's knowledge of irregular verbs is the correct usage rate (CUR), $C(x)$, defined as follows:

(46) $$C(x) = \frac{\text{total number of correct past tense of } x}{\text{total number of past tense of } x}$$

Our data on child performance come from the monograph *Overregularization in Language Acquisition* (Marcus et al. 1992), where four American children (Adam 2;3–5;2, Eve 1;6–2;3, Sarah 2;3–5;1, and Abe 2;5–5;0) were studied, using the longitudinal recordings transcribed in the CHILDES corpus (MacWhinney & Snow 1985).[11] Marcus et al. manually analyzed the transcripts, and hence eliminated the unavoidable ambiguity that may have escaped computerized pattern extractions.[12] The input frequencies of irregular verbs are determined by the present author, based on more than 110,000 adult sentences to which Adam, Eve, Sarah, and Abe were exposed during the recording sessions.

The CURs of all irregular verbs, averaged over all recording sessions, are computed from Marcus et al. (1992: tables A1–A4) and given in (47):

(47) a. Adam: 2446/2491 = 98.2%
b. Eve: 285/309 = 92.2%
c. Sarah: 1717/1780 = 96.5%
d. Abe: 1786/2350 = 76%

The average CUR for the four children is 89.9%. It is clear that there is quite a bit of individual variation among the children. While Adam, Eve, and Sarah used irregular verbs almost perfectly, Abe's performance is markedly worse. Of particular interest is the verb class [-ø & Rime → U], which includes verbs such as *know*, *grow*, *blow*, *fly*, and *throw*. This class posed significant difficulty

[11] Other children studied in the monograph are not included here, because of the relatively small size of their recordings and the lack of longitudinal data.

[12] For example, the past tense of no-change irregular verbs can only be accurately identified from the conversation context.

for all four children. The CURs are 7/16 = 44% (Adam), 0/1 = 0% (Eve), 12/22 = 55% (Sarah), and 28/71 = 39% (Abe). For Adam, Eve, and Sarah, this is the only seriously problematic class. We will explain this peculiar pattern in section 3.3.4.

The WR model learns and organizes irregular verbs on the principle of frequency-sensitive associative memory: the more you hear, the better you remember and the better you retrieve. Hence, $C(x)$ for the WR model is correlated with the frequency of x in past tense form, f_x. In the RC model, the performance of an irregular verb x is determined by two factors: the probability that x is associated with its class S, and the probability f_S of the rule R applying over the default *-d* rule. Hence, $C(x)$ in the RC model is correlated with $f_x \times \sum_{m \in S} f_m$.

3.3.3 *Frequency hierarchy in verb classes*

The first prediction made by the RC model is straightforward:

(48) For two verbs x_1 and x_2 *within* a verb class, $C(x_1) > C(x_2)$ if $f_{x1} > f_{x2}$.

To test this prediction, we have listed some verbs grouped by class in (49), along with their input frequencies estimated from adult speech.[13] To make intra-class comparison, only non-trivial classes are included. Also, to minimize sampling effect, only verbs that were used by children at least twenty times are included in our study (Appendix C gives a complete list of irregular verbs with their frequencies):

(49)	*Verbs grouped by class*	*Input frequency*
a.	[*-t* & Vowel Shortening]	
	lose (80/82=97.6%)	*lost* (63)
	leave (37/39=94.9%)	*left* (53)

[13] Past tense forms that can be unambiguously determined (e.g. *drew*, *took*) were counted by an automated computer search. Ambiguities that arise between past tense and present tense (e.g. *hit*), past participles (e.g. *brought*, *lost*), nouns (e.g. *shot*), and adjectives (e.g. *left*) were eliminated by manually combing through the sentences in which they occurred. Since we are comparing the relative CURs for verbs within a single class, no effort was made to distinguish past tense *put* and *got* from their participle forms, as it is clear that their frequencies thoroughly dominate other members in their respective classes.

b.	[*-t* & Rime → a]	
	catch (132/142=93.0%)	*caught* (36)
	think (119/137=86.9%)	*thought* (363)
	bring (30/36=83.3%)	*brought* (77)
	buy (38/46=82.6%)	*bought* (70)
c.	[-ø & No Change]	
	put (239/251=95.2%)	*put* (2,248)
	hit (79/87=90.8%)	*hit* (66)
	hurt (58/67=86.6%)	*hurt* (25)
	cut (32/45=71.1%)	*cut* (21)
d.	[-ø & Vowel Shortening]	
	shoot (45/48=93.8%)	*shot* (14)
	bite (33/37=89.2%)	*bit* (13)
e.	[-ø & Backing ablaut]	
	get (1269/1323=95.9%)	*got* (1,511)
	take (118/131=90.1%)	*took* (154)
	write (20/27=74.1%)	*wrote* (28)
	win (20/36=55.6%)	*win* (36)
f.	[-ø & Rime → u]	
	know (17/23=73.9%)	*knew* (49)
	throw (11/34=32.4%)	*threw* (28)

(49) strongly confirms the prediction in (48): within a single class, the more frequently a verb is heard, the better its CUR.[14] The 'exception' in class (49b), where *think*, a more frequent verb than *catch*, is used at a lower CUR, is only apparent. It is an averaging effect, as (50) makes clear:

(50)	*Children*	*Verb*	*% correct*
a.	Adam, Eve, & Sarah	*think*	100% (44/44)
		catch	96.5% (110/114)
b.	Abe	*think*	80.6% (75/93)
		catch	78.6% (22/28)

The low averaged CUR of *think* in (49b) is due to a disproportionately large number of uses from Abe. Once individual variations are factored out as in (50), it is clear that *think* is used correctly at a higher frequency than *catch*, as predicted.

[14] The strong frequency–CUR correlation in the class [-ø & Backing ablaut] might not be taken at face value. The sound-changing patterns in this class are not homogeneous as in other classes, but are nevertheless conventionally labeled altogether as 'Backing ablaut'. See also n. 4.

(49) reveals a very important pattern: when verbs are grouped into classes defined by phonological rules, their performance is, virtually without exception, ordered by their input frequencies. This unequivocally points to the conclusion that irregular verbs are organized in (rule-defined) classes. This generalization cannot be stated in theories that do not have verb classes. In fact, the frequency–overregularization correlation is also considered by Marcus et al. (1992: 118), who found that for the nineteen children tested, the correlation efficient is –0.37—significant, but far from perfect. What the WR model shows is that frequency plays an important role in the performance of irregular verbs; what it does not show is the precise manner in which frequency affects performance.

The frequency–performance correlation almost completely breaks down when verbs from *different* classes are considered. To see this, we turn to the second prediction made by the RC model, which reveals more empirical problems for the WR model.

3.3.4 *The free-rider effect*

Recall that the RC model predicts:

(51) For two verbs x_1 and x_2 such that $x_1 \in S_1$, $x_2 \in S_2$ and $f_{x1} = f_{x2}$, $C(x_1) > C(x_2)$ if $f_{S1} > f_{S2}$.

(51) means that the CUR of an irregular verb *x* could be quite high even if it is relatively infrequent, as long as other members of its class *S are* frequently encountered. This 'free ride' is made possible by the rule shared by all members of a class.

Since most high-frequency verbs are used correctly, we direct our attention to verbs in (49) that have the lowest input frequencies: *hurt* (25), *cut* (21), *bite* (13), and *shoot* (14). (We postpone the discussion of *bite* and *shoot* to section 3.3.5 for reasons that will become clear there.) We have also included *blew, grew, flew*, and *drew*, which appeared 5, 7, 14, and 22 times respectively, and belong to the [-ø & Rime → u] class that is problematic for all four children.

Consider the six irregular verbs in (52):

(52) Different performance with comparable frequencies (≤ 25 occurrences)

	Verb class	Verbs	% correct
a.	[-ø & No Change]	*hurt, cut*	80.4% (90/112)
b.	[-ø & Rime → u]	*draw, blow, grow, fly*	35.2% (19/54)

Despite the comparable (and low) input frequencies, the verbs in (52a) and (52b) show a sharp contrast in CUR. This is mysterious under the WR model.

Furthermore, consider the asymmetry between *hurt* and *cut* with *know* and *throw*, the latter of which have considerably higher input frequencies than the former:

(53) Higher performance despite lower frequencies

	Verb class	Verb (frequency)	% correct
a.	[-ø & No Change]	*hurt* (25), *cut* (21)	80.4% (90/112)
b.	[-ø & Rime → u]	*know* (58), *throw* (31)	49.1% (28/57)

Here the verbs in (53a) are used better than those in (53b), despite their lower input frequencies. Again, it is not clear how the WR model accounts for this.

The asymmetries observed in (52) and (53) straightforwardly fall out of the RC model for a simple reason: the rule for (52a) and (53a) has much higher weights than those in (52b) and (53b), the free-rider effect. The first rule applies to the verbs *hurt* and *cut*, which do not change in past tense forms. The rule for this class, namely, [-ø & No Change], is amply represented in the input, including *hit, let, set, cut, put*, etc, which have very high usage frequencies, totaling over 3,000 occurrences. Every occurrence of such verbs increases the weight of the class rule. Hence, *hurt* and *cut* get a free ride, and have a high CUR despite a low absolute frequency. In contrast, verbs in (52b) belong to the [-ø & Rime → u] class (*blow, grow, know, throw, draw*, and *fly*), which totals only 125 occurrences in the input sample. Hence, the weight of the rule [-ø & Rime → u] must be considerably lower than that of [-ø & No Change]: the CUR asymmetry in (52) is thus accounted for.

A closer look at Abe's performance, which is markedly poor across all verb classes, reveals an even more troubling pattern for the WR model. Consider the verbs and their CURs in (54):

(54) Lower performance despite higher frequencies (Abe)

Class	Verbs (frequency)	% correct
suppletion	*go* (567)	0.646 (117/184)
[-ø & umlaut (ʌ → ey)]	*come* (272)	0.263 (20/76)

The verbs in (54) are among the most common words in English and have far higher frequencies than those in (52a). However, for the low-frequency verbs in (52a), Abe has an average CUR of 0.659 (29/44, Marcus et al. 1992: table A8): in fact better than *went* and *came*.

This peculiarity in Abe's performance is readily explained by the RC model. Despite their relatively high frequencies, *go-went* and *come-came* nevertheless 'act alone', for they are in trivial classes. The suppletion case of *go-went* is obvious. *Come-came* belongs to the heterogeneous class [-ø & umlaut], which in fact consists of three subclasses with distinct sound changes: *fall* and *befall*, *hold* and *behold*, and *come* and *become*. Hence, *come* only receives help from *become*, which isn't much: two occurrences in all of the input.[15]

3.3.5 *The effect of phonological regularity: Vowel Shortening*

Consider the following two low-frequency verbs: *shoot* and *bite*, whose past tense forms appeared only 14 and 13 times respectively in more than 110,000 adult sentences. Nevertheless, they are used virtually perfectly—91.8% (78/85)—again in sharp contrast with the performance (40.5%) on the verbs in the [-ø & Rime → u] class (52b).

Past tense formation for both *shoot* and *bite* fall under the rule [-ø & Vowel Shortening]. As remarked in section 3.2.2 and in (42), Vowel Shortening is a pervasive feature of the English language. Furthermore, Myers (1987) and Halle (1998) show, from different perspectives, that Vowel Shortening is essentially free: vowels in closed syllables are automatically shortened under suffixation, resulting from the interaction between universal phonological constraints and language-specific syllabification properties. Given the evidence that (English) children have good grasp of the syllabic structure of their language (e.g. Smith 1973, Macken 1980), and that they perform morphological analysis of words from early on (Clark 1993), learning irregular verbs with Vowel Shortening is considerably simplified; in fact, reduced to learning which suffix (*-t*, *-ø*, or *-d*) is attached. And children are very good at learning suffixes, as we saw when reviewing their agreement morphology acquisition in section 3.2.2.

[15] Abe's performance on the other two umlaut subclasses are not much better: *fall-fell* is used correctly 72 times out of 129 uses, while *fell* appeared 279 times in the adult input, and *hold-held* is used correctly 0 of 4 times, while *held* appeared 11 times in the adult input, although the sample size in the latter case is too small to be truly informative.

In (55), we see that all three classes of Vowel Shortening verbs have very high CURs:

(55) Vowel Shortening under suffixation

	Suffix	Verb (frequency)	% correct
a.	[-t]	*lose-lost* (63)	98% (80/82)
		leave-left (53)	95% (378/39)
b.	[-d]	*say-said* (544)	99% (522/525)
c.	[-ø]	*shoot-shot* (14)	94% (45/48)
		bite-bit (13)	90% (33/37)

All verbs in (55) are used very well, almost irrespective of their individual frequencies, ranging from very frequent ones (*say-said*) to very rare ones (*shoot-shot*, *bite-bit*). Such complete frequency defiance, along with the asymmetries noted in (52), (52b), and (54), strongly point to the reality of class-defining phonological rules in the RC model.

3.4 Analogy, regularity, and rules

3.4.1 *The failure of analogy*

Section 3.3 has identified a major problem with the WR model. The regularity among verbs in a class, expressed in a shared phonological rule in the RC model, is not statable in the WR model.

Perhaps the notion of analogy, built on phonological similarity (of some sort), may duplicate the effect of rules without explicitly assuming them. This is the only way to account for the acquisition data where frequency–performance correlation breaks down. Consider Pinker's discussion on analogy:

> Analogy plays a clear role in language. Children, and adults, occasionally analogize the pattern in one regular verb to a new irregular verb (*write-wrote* → *bite-bote*). They also find it easier to memorize irregular verbs when they are similar to other irregular verbs. The analogizing is a hallmark of connectionist or parallel distributed processing associators; it suggests that human memory might be like a pattern associator. (Pinker 1995: 129)

As an example, Pinker goes on to suggest that rhyme may play a role in pattern association and memorization. For example, since *draw-drew, grow-grew, know-knew*, and *throw-threw* rhyme with each other, memorizing *draw-drew* facilitates the memorization of other irregular verbs, and vice versa. The *bite-bote* type error results from the occasional misuse of the rhyme analogy.

The alert reader might realize at this point that we have already seen empirical evidence that analogy by rhyme cannot be correct. In sections 3.3.4 and 3.3.5 we have compared children's performance on several low-frequency verbs. Of particular interest are verbs like *shoot-shot* and *bite-bit*, which were used very well, and verbs like *grow-grew* and *blow-blew*, which were used very poorly. However, note that the only irregular verb that *bite-bit* rhymes with is *light-lit*, which appeared only once in the more than 110,000 adult sentences sampled. Worse, *shoot-shot* does not rhyme with *any* irregular verb in English. If Pinker were correct in suggesting that rhyme helps irregular verb memorization, one would expect that *drew, grew, threw*, and *knew*, which rhyme with each other and thus help each other in memorization, would have *higher* retrieval success than *shot* and *bit*, which get help from no one. However, this is not the case.

Could some different forms of analogy (other than rhyme) work so that the WR model can be salvaged? One cannot answer this question unless a precise proposal is made. The question of how words are analogous to each other, and how analogy is actually used to facilitate learning, is usually left vague in the literature, under the rubric of the Wittgensteinian 'family resemblance' (e.g. Bybee & Slobin 1982, Bybee & Moder 1983, Pinker 1999). Here there is a methodological point to be made. While there is evidence that some human concepts cluster around fuzzy 'family resemblance' categories (Rosch 1978; but see Fodor 1998), rather than well-defined classical categories, there is no reason to suppose that the lexicon is organized in a similar way. Furthermore, the goal of modern cognitive science is to understand and model mental functions in precise terms. If one were to be content with vague ideas of analogy or association, such as the passage from Pinker & Prince (1994) quoted earlier, the systematic regularities among irregular verbs noted in section 3.3 will simply escape attention: they are revealed only under scrutiny of the empirical data guided by a concrete theoretical model proposed here.

Empirically, the 'fuzziness' in the use of past tense (Bybee & Slobin 1982, Bybee & Moder 1983) in no way shows that the organization of irregular verb phonology is built on 'prototypes' or 'analogy'. Rather, it simply reflects the probabilistic associations between words and rules, and the probabilistic competitions among rules, as the RC model demonstrates.

It seems that in order to capture the class-based frequency hierarchy reported in (49), the WR model must duplicate the class-defining effect of rules with 'analogy', the type-level association based on phonological similarities of verbs (in a class). But analogy works only when the sound similarities among verbs under identical rules/classes are strong enough *and* the sound similarities among verbs under different rules/classes are weak enough. A careful look at the irregular verbs in Appendix B shows this is highly unlikely. For example, verbs in the [-ø & No Change] class, such as *hit, slit, split, quit*, and *bid*, are very similar to those in the [–ø & Lowering ablaut] class, such as *sit* and *spit*, yet the two groups are distinct. Phonological similarity does not give a one-to-one mapping from verbs to classes, and that is why the traditional view in phonology (Chomsky & Halle 1968) treats verb and class association by fiat.

Or, consider the free-rider effect discussed in section 3.3.4, where phonological rules enable low-frequency verbs to be used with high accuracy. In order for the WR model to capture the free-rider effect with analogy, the 'family resemblance' among verbs of all frequencies must be very strong. This again leads one to expect that the learner will also strongly 'analogize' past tense formation to verbs that *do not* belong to the class but nevertheless *do* bear a superficial 'family resemblance' to the class members. For example, *think* may be analogized to *sing* and *ring* to yield *thank* or *thunk*. But children do not do this: about 0.2% in all verb uses are analogical errors (Xu & Pinker 1995).

Once we move beyond the impoverished morphology of English and on to other languages, it becomes immediately obvious that the use of phonological rules in the mental lexicon is inevitable. To take an example from Marcus et al. (1995), noun plurals in German employ five suffixes: *Kind-er* (children), *Wind-e* (winds), *Ochs-en* (oxen), *Daumen-ø* (thumbs; using an empty suffix like the English plural *moose-ø* and past tense *hit-ø*), and *Auto-s* (cars). The authors convincingly argue, using a sort of *Wug*-test with novel German nouns, that despite its low frequency, the *-s* is the default plural suffix, However, it is hard to imagine that German speakers memorize all four classes of irregular plurals—the majority of nouns in the language—on a word-by-word basis, as if each were entirely different from the others. It would also be a massive waste of memory.[16] Furthermore, it is the partial similarity among English irregular verbs that led Pinker and his colleagues to look for family resemblance:[17] four irregular classes of German noun plurals do not show any systematic similarity. Hence, no analogy comes to the rescue. It seems that German learners must sort each irregular noun into its proper class, as suggested by the traditional rule-based view.

[16] This inefficiency of memorization is not dramatic in English, a language with a very small irregular vocabulary.

[17] Which seems no more than a historical accident: see section 3.4.2.

The problem gets worse when we turn to languages with agglutinative morphology such as Turkish, or the so-called 'polysynthetic' languages (Baker 1996). These languages typically have very long 'words' built out of many morphemes, each of which expresses an individual meaning and all of which are glued together by both the morphophonological and the syntactic systems of the language. It is inconceivable that these 'words', which realize millions or billions of morphological feature combinations, are all individually memorized: some sort of combinatorial system must be employed.

This is not to say that analogy plays no role in learning. Mis-irregularization errors such as *bring-brang* in children and adults do seem analogy-based (Prasada & Pinker 1993).[18] However, the role analogy plays in learning must be highly marginal—precisely as marginal as the rarity of analogy errors, 0.2%. This suggests that a very weak effect of phonological analogy can be realized in the verb-to-class linking component of the RC model. As for an overall theory of past tense, it is important to realize, as Pinker & Prince (1988: 127, italics original) remark, that 'a theory that can *only* account for errorful or immature performance, with no account of why the errors are errors or how children mature into adults, is of limited value'. A model that banks on analogy, which can only explain weird past tense errors, misses the major target of the study.

3.4.2 *Partial regularity and history*

Before moving on, let us consider a major objection of proponents of the WR model to the rule-based approach. Since an irregular verb forms past tense by fiat, according to generative

[18] As pointed out to me by Noam Chomsky and Tom Roeper, by far the most frequent pattern in children's weird past tense errors involve verbs with an *-ing* ending such as *bring-brang* (Xu & Pinker 1995: table 2). In addition, *brang* is even acceptable to some speakers. Indeed, errors such as *bite-bote*, cited by Pinker (1995), and many conceivable errors (e.g. *think-thunk* after *sink-sunk*, *hit-hat* after *sit-sat*) were not found. This again suggests that analogy is a very weak influence.

phonology, there is no explanation why verbs like *sting, string, sling, stink, sink, swing,* and *spring* all change *i* to *u* in the past participle and all sound so similar (e.g. Pinker 1999: 102). Pinker's explanation is again based on family resemblance, the sort of fuzzy associations borrowed from connectionist networks. Since verbs are represented as bits and pieces of sound segments (Pinker & Prince 1994, Pinker 1999), the common parts they share are reinforced most often and thus become gravitational attraction for word families, with some prototypes close to the center such as *string-strung* and *sling-slung*, and some on the fringes such as *dig-dug* and *win-won*. But this reasoning seems circular: why are *these* verbs pulled into similarity-based families? As far as one can tell, because they sound similar. Also notice that stem similarity is only partial: the *i-u* family does not include *think*, whose past participle is *thought*, or *blink*, which is altogether regular, and both of them seem closer to the family center than *dig* and *win*. Nowhere does the WR model specify how fuzzy family resemblance actually works to prevent *thunk* and *blunk* from being formed.

The most important reason for this misguided challenge is, partial regularity in verb classes is a result of historical contingencies.

In the RC model, verb classes are defined by rules such as (41), repeated below:

(56) Rule R for verb class S

$$x \xrightarrow{R} y \text{ where } x \in S = \{x_1, x_2, x_3, \ldots\}$$

The members of S are simply listed, and they share the R, which computes the output form, y. One can imagine another kind of rule that is defined in terms of *input*, where the past tense of the verb is entirely predictable from the stem:

(57) Rule R for verb class S

$$x \xrightarrow{R} y \text{ where } x \text{ has property } \pi_S$$

In present-day English, rules like (57) are full of exceptions, at least in the domain of the past tense. However, their regularities were higher further back in history. Even the suppletive verbs,

which may seem arbitrary synchronically, are not necessarily accidents diachronically. In Middle English, for example, *go* somehow replaced the now obsolete *wend*. However, *go* did retain the past tense form, *went*, which belongs to the more regular class that also includes *bend* and *send*. Hence, the suffixation and readjustment rules, synchronically productive, are evidenced diachronically: no irregular verbs are exceptions to *-t*, *-ø*, and *-d* suffixation.

How did such (partial) regularities get lost in history? There are two main factors; see Pinker (1999: ch. 3) for a good discussion. One is purely frequency-based. If an irregular verb is used very infrequently, the learner will not reliably locate it in the appropriate class to which it belongs. We will return to this in section 3.5.9. The other factor falls out of the interaction between irregular rules and changes in other parts of the phonological system. See Pinker (1999: 65) for the history of the now archaic *wrought*. The evolution of irregular verbs is not completely random, therefore, but rather stochastic: sampling effects and other unpredictable changes, such as *go* replacing *went*, interact with predictable UG principles and conventions to produce partial similarities observed in irregular verb classes. The reader is referred to Yang (2002) for a formal model of sound change based on the RC model of learning, and for a detailed discussion of these issues.

3.5 Some purported evidence for the WR model

Pinker (1995) summarizes previous work on the WR model and gives ten arguments in its support. Here we review them one by one, and show that, where they are not factually inaccurate or methodologically flawed, they are handled equally well or better by the RC model.

3.5.1 *Error rate*

How low is it?

Pinker claims that the rate of past tense errors is quite low: the mean rate across twenty-five children is 4.2%, the median only

2.5%. He suggests that this low rate indicates that overregularization is 'the exception, not the rule, representing the occasional breakdown of a system that is built to suppress the error', as in the WR model.

First, it is questionable whether the actual error rate is actually *that* low. In (47), we saw that the error rate averaged over four children is 10.1%. In particular, Abe's error rate is *very* high: about 24% of the irregular verbs were regularized. Also, as is clear from Marcus et al. (1992: table A8), Abe's poor performance is systematic and cuts across all verb classes, and thus is not due to a few particularly bad and very frequent verbs/classes.[19] He even made a considerable number of errors (64/177=36%) in *go-goed*, while all other children used *went* perfectly throughout. Second, by averaging over all irregular verbs, the more problematic but less frequent verbs and classes and the important variations among classes (section 3.3) are lost. For example, all four children performed very badly on the [-ø & Rime → u] class, an error rate of 48.6% (54/111).

Longitudinal trends

Pinker claims that the rate of overregularization, 2.5%, is stable through the preschool years (2–5), and gives Adam's longitudinal overregularization trend, which is indeed quite steady (and low) over time. He concludes that the steady error rate is due to the occasional malfunction of memory retrieval—the exception, not the rule.

There are strong reasons to challenge this claim. First, it seems that Adam is the exception, rather than the rule. Adam's grasp of irregular verbs is in general perfect, the best among the four children we examined; see (47). Second, as already noted in section 3.5.1, averaging over all irregular verbs is likely to obscure longitudinal patterns, which could be observed only in problematic verbs (e.g. the *know-knew* class).

[19] See Maratsos (2000) for a discussion of Abe, in particular why the large set of data from Abe must be taken as seriously as those from other children.

Fortunately, we do have Abe, whose irregular verb performance is, across all verb classes, markedly worse than the other three children. To study Abe's longitudinal development, we have grouped every consecutive fifteen recordings into a period. There are 210 recordings (from 2;4 to 5;0), so we have fourteen periods altogether. We have examined verbs that Abe was particularly bad at: *go, eat, fall, think, came, catch, run*, and the members of the problematic [-ø & Rime → u] class: *throw, grow, know, draw, blow*, and *fly*. The results are are summarized in Table 3.1.

With the exception of period 1, in which Abe only had eighteen opportunities to overregularize (and there was thus a likely sampling effect), his error rate is gradually declining. This shows that children's overregularization at the earliest stage is considerably more significant and systematic than Pinker claims, and cannot be attributed simply to 'exception'.

3.5.2 *The role of input frequency*

Pinker notes that the more frequently an irregular verb is heard, the better the memory retrieval for that verb gets, and the lower the overregularization rate. This claim, while correct for verbs

TABLE 3.1. Abe's longitudinal overregularization for problematic verbs

Period	No. of overregularization	Total no. used	Error rate
1	3	18	0.167
2	14	25	0.560
3	31	50	0.620
4	27	37	0.729
5	10	19	0.526
6	28	56	0.500
7	28	54	0.519
8	7	38	0.184
9	18	52	0.346
10	10	40	0.250
11	4	33	0.121
12	4	23	0.174
13	2	43	0.047
14	3	46	0.065

within a class (section 3.3.3), is in general incorrect. The performance of an irregular verb is determined by two factors: the correct identification of class membership, and the weight of the irregular rule (see sections 3.3.4 and 3.3.5).

3.5.3 The postulation of the -d rule

In the stage which Pinker calls phase 1 (from 2;3 to shortly before 3;0), Adam left many regular verbs unmarked: instead of saying *Yesterday John walked*, the child would say *Yesterday John walk*. Overregularization started in phase 2, as the rate of tensed verbs very rapidly became much higher. Pinker suggests that the two phases are separated by the postulation of the *-d* rule. Although this appears to be a reasonable interpretation, it is problematic when individual variations and other aspects of language acquisition are taken into consideration.

First, individual variations. Pinker (1995) only gives the tendency of regular verb marking for Adam, based on Marcus et al. (1992: 109). However, on Marcus et al. (1992: 109–11) we see that the other three children showed very different patterns. Eve's use of regular verbs was basically in a steady climb from the outset (1;6). Sarah showed quite a bit of fluctuation early on, perhaps due to sampling effect, before gradually settling on an ascent. Abe, whose irregular verbs were marked poorly, nevertheless showed the highest rate of regular verb marking: he started out with about 70% of regular verb marking at 2;5, rising to 100% around 2;10.

Second, the low rate of tense marking in phase 1 may be complicated by the so-called Optional Infinitive (OI) stage, first reported by Weverink (1989). Children learning some but not all languages (including English) go through a stage in which they produce a large amount of nonfinite as well as finite verbs in matrix sentences as well as finite. Although there is no consensus on how OI should be explained, to the extent that the phenomenon is real, it may cause the lack of past tense marking.

Consider an alternative explanation of the rapid increase Pinker noted in the use of inflected verbs. No discontinuity in the *-d* rule is supposed; that is, we assume that the *-d* rule is learned by the child quite early on, perhaps along the lines suggested by Yip & Sussman (1996, 1997). However, during the OI stage, the *-d* rule, which applies to past tense verbs, simply does not apply to the extensively used nonfinite verbs that are allowed by an OI stage competence system. When children leave the OI stage, the *-d* rule consequently becomes applicable.

A good test that may distinguish this position from Pinker's is to turn to a language for which the OI stage does not exist, so that OI is not a confounding factor. Italian and Spanish are such languages, where children reliably inflect verbs for tense (Guasti 1992, Wexler 1998). If the alternative view, that the *-d* rule is available from early on, is correct, we predict that in the acquisition of Italian and Spanish, irregular verbs ought to be overregularized from early on. The late postulation of the *-d* rule in the WR model does not make this prediction. So far we have not checked this prediction.

3.5.4 Gradual improvement

Pinker notes that after the *-d* rule is postulated (but see the previous section for an alternative view), overregularization does not drive out correct use of irregular verbs, but bare forms instead, which are extensively used during phase 1. He cites Adam's performance for support. Adam's average CUR is 0.74 during phase 1, and 0.89 during phase 2. There appears to be no 'real regression, backsliding, or radical reorganization' (1995: 118) in Adam's irregular verb use. This follows if the memory for irregular verbs is getting better.[20]

Gradual improvement is also predicted by the RC model, as weights for class membership and irregular rules can only increase. The gradual improvement in the performance results from the increasing amount of exposure to irregular verbs.

[20] The gradual improvement in Adam's performance seems to contradict Pinker's earlier claim that Adam's error rate is stable (section 3.5.1).

3.5.5 *Children's judgement*

Experiments have been conducted to test children's knowledge of irregular verbs, by presenting them with overregularized verbs and asking them if they sound 'silly'. Children are found to call overregularized verbs silly at above chance level. This finding is claimed to show that children's grammar does judge overregularization as wrong, despite their occasional use of it.

Pinker correctly points out some caveats with such experiments: a child's response might be affected by many factors, and thus is not very reliable. In any case, these findings are hardly surprising: even Abe, the child with by far the worse irregular verb use, had an overall error rate of 24%—far better than chance. In fact, such findings are compatible with any model in which children produce more correct forms than overregularizations at the time when judgements were elicited.

3.5.6 *Anecdotal evidence*

Pinker cites two dialogues (one is given below) between psycholinguists and their children, during which the adults use overregularized verbs to observe the children's reaction. The children are not amused.

Parent: Where's Mommy?
Child: Mommy goed to the store.
Parent: Mommy goed to the store?
Child: NO! (*annoyed*) Daddy, *I* say it that way, not you.

Pinker (1995: 119) suggests that the children, 'at some level in their minds, compute that overregularizations are ungrammatical even if they sometimes use them themselves'.

Whether anecdotal evidence should be taken seriously is of course a concern. Possibly, children do not like to be imitated. In any case, the RC model gives a more direct explanation for observed reactions. Recall that at the presentation of each past verb, the child has probabilistic access to either the special irregular rule

(when applicable) or the default *-d* rule, to generate the expected past tense form from the extracted root. Now if an overregularized form such as *goed* is repeated several times, the chance of a mismatch (i.e. the child generating *went*) is consequently enhanced—the probability of generating *went* at least once in several consecutive tries—much to children's annoyance, it appears.

3.5.7 *Adult overregularization*

Adult do occasionally overregularize. Pinker claims that the rarity entails that adult overregularization is the result of performance, not the result of a grammatical system. However, this is not the only interpretation of adult overregularization: rule-based grammatical system approaches account for the data equally well. Under the RC model, for an irregular verb (e.g. *smite-smote*) that appears very sparsely, the learner may not be sure which class it belongs to, i.e. the probability of class membership association is considerably below 1. Overregularization thus results, even if the weight of the irregular rule for its corresponding class is very close to 1.

Pinker also notes that since memory fades when people get older, more overregularization patterns have been observed during experiments with older people (Ullman et al. 1993). This interesting finding is consistent with every theory that treats the irregulars as different—cognitively, and ultimately neurologically—from the regulars: in the RC model, it is the class membership that is memorized.

3.5.8 *Indecisive verbs*

Adults are unsure about the past tense of certain verbs that they hear infrequently. *Dreamed* or *dreamt*? *Dived* or *dove*? *Leapt* or *leaped*? *Strided* or *strode*?[21]

[21] Some of those forms are doublets, so both forms are heard. As noted in section 3.2.4, they pose a problem for the Absolute Blocking Principle, which the WR model adopts.

Pinker links input frequency to the success of irregular past tense (memory imprint). Again, this correlation is also expected under the RC model: low-frequency verbs give the learner little clue about class membership, and for doublets, the class membership is blurred by the non-trivial frequencies of both forms.

3.5.9 *Irregulars over time*

Pinker cites Joan Bybee's work showing that, of the 33 irregular verbs during the time of Old English, 15 are still irregular in Modern English, with the other 18 lost to the *+ed* rule. The surviving ones had a frequency of 515 uses per million (137/million in past tense), and the regularized ones had a frequency of 21 uses per million (5/million in past tense). The more frequently used irregulars are retained.

The RC model readily accounts for this observation. Suppose that for generation n, all 33 irregular verbs had irregular past tense forms, but some of them are very infrequently used. As a result, generation $n + 1$ will be unsure about the class membership of the infrequent irregulars, for reasons discussed in section 3.5.8, and will regularize them sometimes. Consequently, generation $n + 2$ will be even less sure and will produce more regularized forms. Eventually, when the irregular forms drop into nonexistence, such verbs will have lost their irregular past tense forever. Thus, the loss of irregularity is a result of sampling effects and competition learning over time. See Yang (2002) for a model that formalizes this process.

3.5.10 *Corpus statistics*

Based on the statistics from modern English text corpora, Pinker found that the top ten most frequently used verbs are all irregular verbs, and that 982 of the 1,000 least frequently used are regular verbs. He reasons that this pattern is predicted, since the survival of irregular verbs against children and adults' overregularization is only ensured by high frequency of use. This is certainly correct,

but is also obviously compatible with the RC model, following the discussion in 3.5.8 and 3.5.9.

3.6 Conclusion

We have proposed a rule competition model for the acquisition of past tense in English. A list of irregular rules, defined over classes of irregular verbs, compete with the default *-d* rule for past tense inflection. Hence, the learning of an irregular verb is determined by the probability with which the verb is associated with the corresponding irregular rule, as well as the probability of the rule applying over the default *-d* rule. We have also given justifications for, and explored the consequences of, a stochastic and learning-theoretic version of the Blocking Principle.

The RC model is completely general, and applicable to the acquisition of phonology in other languages. Complemented by the Yip–Sussman model of rule learning, our model makes very precise predictions about verb learning: any two verbs can be directly compared (45), based on quantifiable frequency measures drawn from linguistic corpora. Such quantitative predictions are strongly confirmed by the acquisition data. We view the findings here as a strong challenge to any phonological theory that rejects rules.

Scrutiny over past tense 'errors' revealed much about the organization and learning of phonology. In Chapter 4, we turn to their syntactic counterparts.

Appendix B: The rule system for English past tense

This list is loosely based on Halle & Mohanan (1985: appendix) and Pinker & Prince (1988: appendix). Very rare verbs are not listed.

Suppletion

go, be

-t *suffixation*

- No Change
  burn, learn, dwell, spell, smell, spill, spoil
- Deletion
  bent, send, spend, lent, build
- Vowel Shortening
  lose, deal, feel, kneel, mean, dream, keep, leap, sleep, leave
- Rime → *a*
  buy, bring, catch, seek, teach, think

-ø *suffixation*

- No Change
  hit, slit, split, quit, spit, bid, rid, forbid, spread, wed, let, set, upset, wet, cut, shut, put, burst, cast, cost, thrust, hurt
- Vowel Shortening
  bleed, breed, feed, lead, read, plead, meet

-d *suffixation*

- Vowel Shortening
  flee, say
- Consonant
  have, make
- ablaut
  sell, tell
- No Change (default)
  regular verbs

Appendix C: Overregularization errors in children

Irregular verbs are listed by classes; in the text, only verbs with 25 or more occurrences are listed. The counts are averaged over four children. All raw counts from Marcus et al. (1992).

- [-*t* & Vowel Shortening]
  lose 80/82, *feel* 5/18, *mean* 4/5, *keep* 2/2, *sleep* 3/6, *leave* 37/39
- [-*t* & Rime → *a*]
  buy 38/46, *bring* 30/36, *catch* 132/142, *teach* 8/9, *think* 119/137
- [-ø & No Change]
  hide, slide, bite, light
  shoot
- Lowering ablaut
  sit, spit, drink, begin, ring, shrink, sing, sink, spring, swim
  eat, lie
  choose
- Backing ablaut
  I → ʌ *fling, sling, sting, string, stick, dig, win*
  ay → aw *bind, find, grind, wind*
  ay → ow *rise, arise, write, ride, drive, strive, dive*
  ey → u *take, shake*
  er → or *bear, swear, tear, wear*
  iy → ow *freeze, speak, steal, weave*
  ɛ → *a* *get, forget*
- umlaut
  fall, befall
  hold, behold
  come, become
- Vowel → u
  blow, grow, know, throw, draw, withdraw, fly, slay
  hit 79/87, *cut* 32/45, *shut* 4/4, *put* 239/251, *hurt* 58/67
- [-ø & Vowel Shortening]
  feed 0/1, *read* 1/2, *hide* 4/5, *bite* 33/37, *shoot* 45/48
- [-ø & Lowering ablaut]
  sing 3/4, *drink* 9/15, *swim* 0/3, *sit* 5/7, *spit* 0/3
  eat 117/137
- [-ø & Backing ablaut]
  stick 5/10, *dig* 2/5, *win* 20/36
  ride 7/8, *drive* 6/12
  take 118/131, *shake* 4/4
  get 1269/1323, *forget* 142/142

- [-ø & umlaut]
  fall 266/334
  hold 0/5
  come 109/174
- [-ø & Rime → u]
  blow 5/15, *grow* 4/12, *know* 17/23, *throw* 11/34, *draw* 2/12, *fly* 8/15
- [-*d* & Vowel Shortening]
  say 522/525

4

Grammar Competition in Children's Syntax

> Phylogenesis is the mechanical cause of ontogenesis. The connection between them is not of an external or superficial, but of a profound, intrinsic, and causal nature.
>
> Ernst Hackel, *Ontogeny and Phylogeny* (Gould 1977: 78)

Hackel's proposition that 'ontogeny recapitulates phylogeny', which has been drifting in and out of fashion in biology, may well be vindicated in the ontogeny of human language, with a twist. If language is delimited in the finite space of Universal Grammar, its ontogeny might well recapitulate its scope and variations as the child gradually settles on one out of the many possibilities. This is exactly what the variational model leads one to expect, and the present chapter documents evidence to this end.

The variational model also serves another important purpose. If we survey the field of language acquisition, we cannot fail to notice an unfortunate gap between learnability studies and developmental studies. As far as we know, there is presently no formal model that directly explains developmental findings, nor any rigorous proposal of how the child attains and traverses 'stages' described in developmental literature. The variational model intends to fill this gap.

The variational model makes two general predictions about child language development:

(58) a. Other things being equal, the rate of development is determined by the penalty probabilities of competing grammars; cf. (25).
b. As the target grammar gradually rises to dominance, the child entertains coexisting grammars, which ought to be reflected in the non-uniformity and inconsistency in its language.