Journal of Experimental Psychology:
Learning, Memory, and Cognition
2009, Vol. 35, No. 4, 849–866

0278-7393/09/$12.00 DOI: 10.1037/a0015286

Can Imageability Help Us Draw the Line Between Storage and Composition?

Elizabeth L. Prado
Georgetown University and Lancaster University

Michael T. Ullman
Georgetown University

Language requires both storage and composition. However, exactly what is retrieved from memory and what is assembled remains controversial, especially for inflected words. Here, "imageability effects" is introduced as a new diagnostic of storage and a complement to frequency effects. In 2 studies of past-tense morphology, more reliable imageability and frequency effects were found on irregulars than on regulars. An interaction with sex was also observed: Males but not females showed more reliable frequency and imageability effects for irregulars than for regulars; females but not males showed signs of storage for regulars, particularly among higher frequency forms. Overall, the findings validate imageability effects as another diagnostic of storage and suggest that the line between storage and computation is not a simple function; rather, it depends upon the interplay of both item- and subject-specific factors.

Keywords: imageability, storage, composition, inflectional morphology, sex differences

Supplemental materials: http://dx.doi.org/10.1037/a0015286.supp

Knowledge of language requires both knowing lexical items and knowing how those items can be combined to create complex linguistic forms. Although it is clear that language involves both storage and composition, the question of what is memorized and what is assembled as we speak (or disassembled as we listen) is a topic of ongoing research. This broad issue has been investigated across many different aspects of language with a variety of methodological approaches. Here we examine the issue in inflectional morphology using a novel experimental approach.

Various theoretical claims have been made with respect to the storage and computation of inflected forms. On one view, all previously encountered inflected words are stored as whole forms in memory, and new forms (e.g., the novel past tenses *spling–splang, plag–plagged*) are generalized over existing memory traces (Bybee, 1995). Similarly, connectionist "single-mechanism" models assume that all inflected forms, whether previously encountered or new, are represented and processed in a distributed associative memory (Elman, 1996; Plunkett & Marchman, 1993; Rumelhart & McClelland, 1986). According to some single-mechanism models, forms that follow a common and consistent inflectional pattern (e.g., *walk–walked, balk–balked, stalk–stalked*) depend largely on phonological representations in this memory, whereas forms that do not (e.g., *dig–dug, think–thought*) rely to a greater degree on semantic representations (Joanisse & Seidenberg, 1999; McClelland & Patterson, 2002).

In contrast, "dual-system" theories claim that although some or even many inflected forms are stored, others are composed or decomposed by rule-governed processes. In particular, these theories assume that regularly inflected forms (i.e., those that follow a default rule) can be assembled and that irregular forms (i.e., those that cannot be fully specified by a default rule) depend on stored representations in the lexicon. For example, English regular past-tense or plural forms, whose stems take an *-ed* or *-s* suffix (e.g., *walked, played, rats, frogs*), may be composed, whereas other past-tense or plural forms—that is, irregulars—are memorized (e.g., *dug, sprang, mice*, *teeth*).

Although all dual-system theories claim that irregulars depend on some sort of memorized representations, the exact nature of these representations varies across different theories. Traditionally, it has been assumed that irregulars are stored as whole words in rote memory (e.g., Kuczaj, 1977). More recently, it has been argued that irregulars are represented in an associative memory in a distributed but structured manner (Pinker, 1999; Pinker & Ullman, 2002). On a third view, irregular verbs are linked in a rote memory to specific affixes (e.g., *keep* + *-t, dig* + *-0*) and thereby allow the combination of the verb stem and the particular affix, as well as the application of word-specific "stem-readjustment" rules (e.g., *i–a* in *spring–sprang*; Halle & Marantz, 1993).

Dual-system theories also vary with respect to their claims about regulars. According to some theories, regulars are always com-

Elizabeth L. Prado, Department of Neuroscience, Georgetown University, and Department of Psychology, Lancaster University; Michael T. Ullman, Department of Neuroscience, Georgetown University.

Additional materials are available on the Web at http://dx.doi.org/10.1037/a0015286.supp

Support for this project was provided to Michael T. Ullman by a McDonnell–Pew Grant in Cognitive Neuroscience, National Science Foundation Grant SBR-9905273, and National Institutes of Health Grants R01 MH58189 and R01 HD049347. We thank Claudia Brovetto, Ivy Estabrooke, Joshua Hartshorne, and Roumyana Pancheva for their early contributions to this project; Matthew Walenski, Amy Knight, and Robbin Miranda for helpful comments; and Christopher Maloof for statistical advice and analyses.

Correspondence concerning this article and reprint requests should be addressed to Michael T. Ullman, Department of Neuroscience, Georgetown University, Box 571464, Washington, DC 20057-1464. E-mail: michael@georgetown.edu

posed (Marslen-Wilson & Tyler, 1998; Taft, 1979; Tyler et al., 2002). Other theories posit that although regulars can be assembled by rule-governed processes, they can instead be stored and indeed often are. On one such view, all previously encountered regulars are memorized and only new words or pseudowords undergo rule-based computation (Caramazza, Laudanna, & Romani, 1988). A less radical position holds that previously encountered regulars may either be stored or be composed and that the likelihood of their storage depends on various factors. Thus it has been claimed that higher frequency regular forms are more likely to be stored (Alegre & Gordon, 1999; Baayen, Dijkstra, & Schreuder, 1997; Lehtonen, Niska, Wande, Niemi, & Laine, 2006; Meunier & Segui, 1999; Stemberger & MacWhinney, 1986), as are "inconsistent" regulars, such as *glide,* whose stems are phonologically similar to the stems of irregulars (storage of these regulars prevents their irregularization by analogy to these same irregulars; e.g., *glide–glode* or *glide–glid* on analogy to *ride–rod*e or *hide–hid*) (Ullman, 1993, 2001a). It has also been suggested that individual or group differences may affect the storage of regulars, even those that are "consistent" (i.e., whose stems are not phonologically similar to the stems of irregulars; Ullman, 2004, 2005a). For example, it has been posited that females, who appear to have superior verbal memories as compared to males (Kimura, 1999; Kramer, Delis, & Daniel, 1988; Maitland, Herlitz, Nyberg, Backman, & Nilsson, 2004; Trahan & Quintana, 1990), should also be more likely than males to store regularly inflected forms (Hartshorne & Ullman, 2006; Ullman, 2004, 2005a; Ullman, Miranda, & Travers, 2008).

Numerous studies have examined these and related claims with a wide variety of psycholinguistic and neurolinguistic methods (for some reviews, see Clahsen, 1999; Lavric, Pizzagalli, Forstmeier, & Rippon, 2001; Marslen-Wilson & Tyler, 1998; McClelland & Patterson, 2002; Pinker, 1999; Pinker & Ullman, 2002; Ullman, 2001b; Ullman et al., 2005). Although certain patterns have begun to emerge from the data, the conflicting claims discussed above have yet to be completely resolved. This lack of resolution is at least partly due to the fact that a number of these issues have not been adequately addressed with appropriate methodologies. For example, whereas many neurolinguistic techniques are well suited for probing for the existence of distinct neural systems, they cannot easily be employed for testing the representational and computational bases of inflected forms. Even psycholinguistic methods that are particularly well suited for addressing these issues have generally been used to examine the basic regular/irregular distinction rather than more subtle distinctions, such as whether and why some regulars may be rule products and others are stored. Additionally, previous studies might not have taken into account a variety of explanatory or confounding factors, and this would increase the likelihood of false interpretation.

Here we have attempted to address some of these problems, using a novel experimental approach. In two independent studies, we have combined two complementary methods, one new (examining "imageability effects," explained just below) and one well established (examining frequency effects), to investigate the influence and interactions of several factors on the storage of inflected forms: the factors of regularity (irregular vs. consistent regular), sex (females vs. males), and frequency (high- vs. low-frequency regulars), as well as imageability itself (are highly imageable regulars more likely to be stored?). Across the two studies, we examined acceptability ratings of inflected forms and response times for the production of inflected forms as dependent variables. Finally, we attempted to account for a wide range of other variables (in addition to frequency and imageability) that could potentially confound the experimental results.

"Imageability effects" are premised on evidence suggesting that lexical items that are more easily imaged—that is, for which a visual mental image can be more easily formed—are also more easily memorized and accessed. Such facilitation has been shown in a variety of single-word tasks, including studies of cued recall (paired associate learning; Paivio, 1963, 1965; Yarmey & O'Neil, 1969), free recall of word lists (i.e., list learning; Kennet, McGuire, Willis, & Schaie, 2000; Mellet, Tzourio, Denis, & Mazoyer, 1998; Paivio, 1967), reading out loud (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Strain, Patterson, & Seidenberg, 1995), and lexical decision (Balota et al., 2004; James, 1975; Kounios & Holcomb, 1994; Kroll & Merves, 1986). Thus, stored forms, including memorized inflected forms, may be expected to show imageability effects (i.e., correlations between their imageability and measures of their accessibility). For example, if irregular past-tense forms are memorized, those of more imageable verbs should be more easily retrieved. In contrast, forms that are composed rather than retrieved should not show imageability effects, at least when controlling for access to their stored parts. For example, if regular past-tense forms are assembled, one would not expect to find that more imageable regular past tenses are processed more easily. Thus, because imageability effects are expected for stored forms but not for rule products, imageability may be used as a diagnostic for storage.

The logic of testing for imageability effects is essentially the same as that underlying the widely used method of probing for frequency effects. This latter method is based on the fact that lexical forms that are encountered more frequently (i.e., that have a higher frequency in the language) are also more easily accessed in memory (Forster & Chambers, 1973; Frederiksen & Kroll, 1976; Rubenstein, Garfield, & Milliken, 1970). Therefore, stored forms, including inflected ones, should show frequency effects (i.e., higher frequency forms should be accessed more successfully or faster), whereas composed forms should not show this relation, at least once stem access is held constant. Thus frequency can also be used as a diagnostic of storage and retrieval.

However, if both frequency and imageability affect lexical access, both must be taken into account. If only one is accounted for, any observed lack of frequency or imageability effects on stored forms might actually be due to the other factor, in particular if there are any interactions between imageability and frequency. Indeed, evidence suggests that imageability effects may be stronger among low-frequency forms (James, 1975; Strain et al., 1995; Strain, Patterson, & Seidenberg, 2002). Thus, examining imageability effects without including frequency as a factor could potentially lead to erroneous conclusions of composition, for example, if frequency but not imageability predicts processing measures in a given data set. Similarly, it is critical to take into account other factors that could potentially affect measures of lexical access, such as measures of phonological similarity or word length, as is done in the studies reported here.

The two studies presented here examined both imageability and frequency effects, as well as their interactions, on two dependent measures (acceptability ratings and past-tense production response times) in order to probe the influence of several factors hypothesized to affect the storage of inflected forms (regularity, sex, frequency, and imageability). They also accounted for an array of other potential explanatory or confounding variables (e.g., measures of phonological neighborhoods). This comprehensive approach not only generates useful data that are expected to elucidate the representation of inflected forms, and must be taken into account by any theory, but additionally provides the means to tease apart aspects of the specific theoretical claims described above. These claims lead to a number of testable predictions.

First, given that all theoretical perspectives assume that some sort of representation of irregulars is stored, presumably all theories expect that irregularly inflected forms should show both frequency and imageability effects, potentially with interactions between them. Imageability effects on irregulars would fit particularly well with connectionist models in which the processing of irregulars depends largely on semantic nodes (e.g., Joanisse & Seidenberg, 1999). However, it is important to point out that because imageability is a factor that has not been previously considered in models of inflectional morphology, no model has explicitly attempted to take it into account. Indeed, it is not clear how some models, such as those that posit only memorized links between stem and affix (Halle & Marantz, 1993), would account for imageability effects on these forms (for a discussion of how such models address frequency effects on irregulars, see Embick & Marantz, 2005).

In contrast, quite a diversity of outcomes is predicted by the different theories for regulars. According to dual-system views, regularly inflected forms that are rule products should show neither frequency nor imageability effects, whereas those that are retrieved from memory (e.g., high-frequency regulars, or perhaps regulars in females, according to some theories) may be expected to show one or both effects. In contrast, according to perspectives that deny all composition in the processing of inflected forms (Bybee, 1995; Joanisse & Seidenberg, 1999; McClelland & Patterson, 2002), all regulars are stored, and thus all types of regulars may be expected to show effects in all subject groups.

However, as with irregulars, whether or not one actually predicts imageability or even frequency effects on regulars that are stored depends on the particular type of representation that is claimed for them. For example, a lack of either effect on regulars may be consistent with connectionist perspectives. In particular, it has been argued that because regulars have similar stem–past phonological transformations, the general phonological pattern is learned in associative memory, leaving little or no influence to individual word frequencies (Daugherty & Seidenberg, 1992; Seidenberg, 1992). It has been suggested that in such models imageability effects should be found only among low-frequency irregulars, as these depend on semantic representations that are influenced by imageability; in contrast, neither higher frequency irregulars nor any regulars should show such effects, as these forms are posited to depend largely on phonological representations (Strain et al., 1995). However, to our knowledge these connectionist claims of imageability have been made only for the domain of reading- aloud words with irregular or regular spelling-to-sound correspondences (e.g., *yacht* vs. *tray*), and it is not yet clear whether or how the argument might also apply to the domain of regular and irregular inflectional morphology (though other generalizations between findings from connectionist models of reading aloud to inflectional morphology have been made; McClelland & Patterson, 2002).

Connectionist predictions of sex differences in inflectional morphology also remain unclear. If females have better verbal memories than have males (see above), connectionist models might reasonably predict weaker rather than stronger frequency and imageability effects for regulars in females as compared to males, as females should be more successful at learning the general regular pattern and should thus rely less on the characteristics of individual words. It might also be argued that if females have an advantage at semantic memory as compared to males (which would follow, for example, from the hypothesis that females have an advantage at declarative memory; Hartshorne & Ullman, 2006; Ullman, 2004; Ullman et al., 2008), females may rely more than males on semantic as compared to phonological representations (following the type of connectionist model proposed by Joanisse & Seidenberg, 1999; McClelland & Patterson, 2002). In such a case, females might be expected to show stronger imageability effects than would males, on both regulars and irregulars.

However, it is important to emphasize that until simulations are actually carried out with specific connectionist models regarding the issues at hand (i.e., examining the influence of frequency, imageability, and sex on regular and irregular English past-tense forms), their outcomes will remain unclear. Thus, the present study is designed to generate data that not only can distinguish between existing models but additionally will help to constrain future models and theories.

In the following sections we first explain the measures for the two independent variables (imageability and frequency). Next we present the methods and results for each of the two studies. Finally, we summarize the results and discuss their significance and implications.

Method: Independent Variables

Here we present the measures used for the two independent variables in the two studies: imageability ratings and word frequencies. The methods specific to the dependent variable in each study are described below in sections dedicated to each study.

Imageability Ratings

Subjects. Thirty subjects—15 male and 15 female—performed an imageability-rating task. Subjects gave informed consent and were paid for their participation. All subjects were right-handed native English speakers. None of the subjects were very fluent in any second language based on self-rated fluency. The subjects ranged in age from 16 to 50 years (M = 26.33 years, SD = 9.90) and in education from 10 to 18 years (M = 15.33 years, SD = 2.01).

Materials. Imageability ratings were gathered for 436 verb stems (and other word forms not analyzed or discussed here, including nouns and adjectives). We focused on imageability ratings of word stems rather than of inflected forms for two reasons. First, because inflectional morphology does not change the basic meaning of a word, the imageability of a given word is not expected to vary significantly between the stem and its various inflected forms. Second, based on pilot testing, subjects seemed more uncertain about rating the imageability of inflected forms than of stem forms, particularly for verbs (e.g., *walked* vs. *walk*).

Because this uncertainty could lead to greater variability on the ratings of inflected forms, we chose to use ratings on stem forms.

The ratings task was presented in paper-and-pencil format. Subjects were instructed to circle the number that corresponded to their imageability rating for each word, with 1 representing low imageability and 5 representing high imageability. The verb stems and other forms were presented in separate lists, with distinct instructions and examples given before each list, in which the grammatical category (e.g., verb or noun) of the words in the list was emphasized. All subjects received the lists in the same order (with the verb stems always rated first), but the order of items within each list was reversed for half of the males and half of the females.

Mean imageability ratings were computed for each item over all subjects. For 2 subjects (1 male, 1 female), mean verb imageability ratings (i.e., over all verbs for those subjects) were more than two standard deviations from the overall verb mean (over all verbs and subjects); thus, the item means were recalculated without these subjects' ratings.

Imageability ratings did not differ between the male and female subjects (14 males and 14 females) in either study, computed across all the regular and irregular items used in the analyses in each study: Study 1, male $M = 3.60$, female $M = 3.60$, $t(26) = 0.04$, $p = .97$, F test for equality of variance, $F(13, 13) = 1.27$, $p = .68$; Study 2, male $M = 3.53$, female $M = 3.48$, $t(26) = 0.31$, $p = .76$, $F(13, 13) = 1.02$, $p = .97$. Additionally, the regular and irregular items did not differ in their imageability ratings in either study (reported below in the *Method* section of each study).

Validity of the imageability ratings: Correlations with published norms. In order to test the validity of these imageability ratings, we performed correlations between these ratings and previously published imageability ratings of the same words. The overlapping subset of 125 verbs common to the verb imageability ratings gathered here and the imageability ratings published by Chiarello, Shears, and Lund (1999) correlated strongly with each other, $r(123) = .781$, $p < .001$. Similarly, a significant correlation was found in the overlapping subset (196 verbs) between the imageability ratings published by Toglia and Battig (1978) and our verb ratings, $r(194) = .516$, $p < .001$. These correlations validate the imageability ratings gathered here.

Frequency Counts

Two relative frequency counts were used in combination for all frequency analyses: (a) the Francis and Kucera counts (Francis & Kucera, 1982), derived from 1 million words of text, drawn from a variety of sources (henceforth "FK"); and (b) a frequency count extracted by a stochastic part-of-speech analyzer from 44 million words of unedited Associated Press (henceforth "AP") news wires between February and December 1988 (Church, 1988; Ullman, 1999). The combined frequency was calculated as the natural logarithm of the sum of the raw FK and AP counts, first augmented by 1 to avoid the undefined ln(0).

Study 1

Study 1 investigated the effect of the two independent variables of interest (imageability and frequency) on acceptability ratings of inflected past-tense forms and their stems. Previous studies have shown that lexical factors, such as frequency, do in fact predict acceptability ratings of irregular but not regular past-tense forms (Ullman, 1993, 1999) and thus suggest that acceptability ratings constitute an appropriate dependent measure for the investigation of the storage of inflected forms.

Method

Acceptability ratings of 856 stems of existing verbs and of their inflected past-tense forms were previously acquired (Ullman, 1993). Only monosyllabic verbs are analyzed and discussed here. The irregular verbs were drawn from those listed in Pinker and Prince (1988). Any irregular verb whose regularized past-tense form had an average acceptability rating of greater than 3.5 in Ullman (1993) was considered a doublet verb (e.g., *dive–dove/dived*) and was excluded. Also excluded were no-change verbs, whose past-tense forms do not differ from their stems (e.g., *hit–hit*). Any regular verb whose stem was phonologically similar to any irregular verb (i.e., in having an identical rime in the stem; e.g., *glide*, cf. *ride, hide*) was considered to be an "inconsistent" regular (see above) and was excluded from analysis. Based on these criteria, a total of 89 irregular verbs and 48 consistent regular verbs—the same set of verbs reported in Ullman (1999)—was analyzed. Because only consistent regulars are analyzed here in Study 1 as well as in Study 2, reported below, in the remainder of the text the term *regular* will be used to refer to consistent regulars, unless otherwise specified.

The acceptability ratings were obtained with paper-and-pencil questionnaires. Subjects rated forms from 1 (*least acceptable*) to 7 (*most acceptable*). Each verb form (stem or past tense) that was presented in the context of an appropriate sentence for that form was rated by 32 subjects.[1] The mean rating over subjects constituted the dependent variable.[2] For further details, see Ullman (1993, 1999).

Comparisons between the regular and irregular verbs showed that they did not differ on their imageability ratings: regulars, range = 2.05–4.54, $M = 3.62$; irregulars, range = 1.89–4.64, $M = 3.57$; $t(135) = 0.44$, $p = .66$. However, they differed significantly on both stem frequency (regulars, range = 0.69–9.28,

[1] Some examples may serve to illustrate the types of sentences in which the stem and past-tense forms were presented: "Whenever we misbehave, they *wring* their hands in frustration" (mean acceptability rating = 6.34); "The members of that club *lose* an average of 1,000 dollars a year" (7.00); "The saber-toothed tiger *sprang* onto the back of the giant sloth" (6.06); "Ernest Hemingway *shot* the charging lion right between the eyes" (7.00).

[2] Because ratings were gathered on a large number of verb forms, the full questionnaires were broken down into subquestionnaires. Each subquestionnaire contained both the stem and past-tense sentences for a given verb. Each subject completed one or more subquestionnaires. Thus, although every verb form was rated by 32 subjects, it was not the case that all verb forms were rated by the same 32 subjects (see Ullman 1993, 1999). Because we no longer had the information regarding which subjects had completed which subquestionnaires (the data were acquired more than 10 years previously), the analyses in Study 1 were performed using a traditional regression technique (i.e., with a general linear model) on the mean acceptability ratings (i.e., averaged over subjects) rather than a hierarchical regression (i.e., with a hierarchical linear model, or HLM; see Study 2 for details), because HLM crucially incorporates subject identity in its estimation of error.

M = 4.78; irregulars, range = 2.64–11.04, M = 7.51), $t(135)$ = 7.93, $p < .0001$, and past-tense frequency (regulars, range = 0.00–9.07, M = 4.21; irregulars, range = 1.09–13.07, M = 6.69), $t(135)$ = 6.37, $p < .0001$. Therefore, as we will see below, relevant analyses were performed both on the entire set of regular and irregular verbs and on a subset of frequency-matched regular and irregular items.

Stem Acceptability

Monomorphemic stem forms (e.g., *walk*, *dig*) are necessarily stored, as they represent arbitrary mappings between form and meaning. If acceptability ratings indeed reflect the strength of memory traces of stored forms (Ullman, 1993, 1999), the ratings of both regular and irregular verb stems would be expected to show imageability as well as frequency effects. The demonstration of these effects would validate imageability as another diagnostic measure (complementing frequency) of the storage of inflected forms.

Imageability and stem (unmarked form) frequency were entered into a multiple linear regression model, with stem acceptability as the dependent variable. Within this model, imageability and frequency coefficients were generated separately for the regular and irregular verbs.

As expected, both imageability and frequency significantly predicted stem acceptability for both regular verbs and irregular verbs (see Figure 1).[3] That is, both imageability and frequency effects were found for regular and irregular verb stems. Note that because the Frequency × Imageability interaction over both verb types was nonsignificant, B = 0.002, $t(129)$ = 0.18, p = .860, and this interaction term was not different between the verb types, B = 0.044, $t(129)$ = 1.55, p = .123, it was not included in the final model (in all regression models, B represents the unstandardized regression coefficient, indicating the slope of the regression line; t represents the t statistic of the comparison between B and zero; and p represents the statistical significance of this comparison; degrees of freedom are estimated using a Satterthwaite approximation). Here and elsewhere, the Frequency × Imageability interaction term was only included in the final model if in the initial model the interaction term over both regulars and irregulars yielded a coefficient that was significant at the $p < .10$ level. The interaction term would also have been included if it differed significantly ($p < .10$) between the two verb types and if the term was significant for at least one of them; however, this was not the case in any model.

We examined certain statistical factors that could potentially account for this pattern of effects. First, visual examination of the graphs suggested three potential outliers that could have undue influence on the estimates of the regression. However, removal of these items yielded the same pattern of significance for both imageability and frequency main effects as with the items included.[4] Second, visual comparison of the plots of the standardized residuals for regulars and irregulars indicated potential heteroskedasticity. However, the model with separate covariance structures for regular and irregular verbs also yielded the same pattern of main effects.[5] In the rest of the regression models reported in this article (i.e., the remaining models reported for Study 1 and all models in Study 2), no outliers or serious heteroskedasticity was detected.

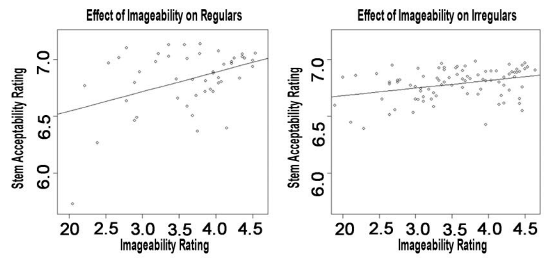
Note that although a significant difference in imageability effects between regulars and irregulars was observed in the original analysis (p = .040; see Figure 1), no regular/irregular differences in imageability or frequency effects were observed either in the model without the potential outliers, imageability, B = 0.031, $t(129)$ = 0.68, p = .500; frequency, B = 0.012, $t(129)$ = 0.81, p = .421, or in the model with separate covariance structures for regular and irregular verbs, imageability, B = 0.105, $t(57.6)$ = 1.72, p = .090; frequency, B = 0.027, $t(64.3)$ = 1.37, p = .176. This suggests a lack of reliable differences in these effects between the two verb types.

Thus, imageability and stem frequency independently predicted stem acceptability. That is, higher stem frequency and higher imageability both led to higher stem acceptability ratings. This suggests that imageability is likely to be a useful tool for investigating the storage and representation of linguistic forms more generally, including the past-tense inflected verb forms examined in this study.

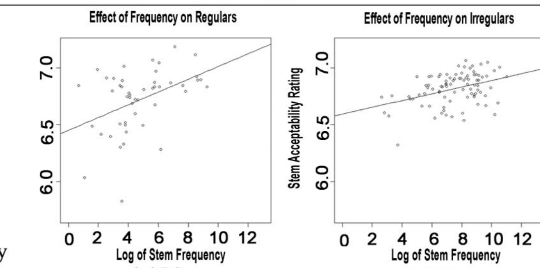
[3] In all scatterplots, the plotted line represents the prediction of the model with regard to one of the two independent variables (frequency or imageability), with all other variables (i.e., the other independent variable and all covariates) held constant at their mean values over the data set. Each point on the graph represents one item (verb form), with its acceptability rating (in Study 1) or log response time (RT; in Study 2) calculated as follows: For each observation of the item in the given data set (e.g., for the RT of *dug* for a given subject in Study 2), subtract, from its acceptability rating or log RT, the estimated effect of every other independent variable, all covariates, and the Frequency × Imageability interaction (if the interaction term is included in the model). Each of these estimated effects is computed by multiplying the coefficient for that variable (e.g., for the same-class neighborhood measure for *dug*) by D, where D represents the difference between the mean value of that variable (over all items) and that particular item's value for that variable. This adjusted dependent-variable value is then plotted (in the case of log RT, the value of each item is adjusted for each subject, and then the mean of the adjusted values for this item is calculated over all subjects).

[4] The potential outliers are the point in the lower left corner of the graph "Effect of Imageability on Regulars," representing the verb *pore*, and the two points in the lower left corner of the graph "Effect of Frequency on Regulars," representing the verbs *pore* and *gore* (see Figure 1). Removal of *pore* and *gore* from the model yielded the same pattern of significance for imageability and frequency effects as with these items included: for regulars, imageability, B = 0.098, $t(129)$ = 2.57, p = .012; stem frequency, B = 0.042, $t(129)$ = 3.65, p = .0004; for irregulars, imageability, B = 0.067, $t(129)$ = 2.58, p = .011; stem frequency, B = 0.029, $t(129)$ = 2.99, p = .003. Note that this analysis excluding potential outliers included the same terms as the corresponding analysis with all items (i.e., the same covariates and no Frequency × Imageability interaction term).

[5] A model specification test (White, 1980) confirmed our visual impression of heteroskedasticity (i.e., heterogeneity of variance in the dependent variable along different values of one or more of the independent variables): $\chi^2(11, N = 137)$= 25.38, p = .008. The model with separate covariance structures for regular and irregular verbs yielded the same pattern of significance as the original model: for regulars, imageability, B = 0.172, $t(45)$ = 3.01, p = .004; stem frequency, B = 0.056, $t(45)$ = 3.18, p = .003; for irregulars, imageability, B = 0.067, $t(86)$ = 3.16, p = .002; stem frequency, B = 0.029, $t(86)$ = 3.67, $p < .001$. These results indicated that heteroskedasticity did not account for the observed pattern of effects.



	Regular Verb Stems	Irregular Verb Stems	Difference: Regular vs. Irregular
Imageability			
B	0.172	0.067	0.105
t(131)	4.19	2.25	2.07
p	<.0001	.026	.040



	Regular Verb Stems	Irregular Verb Stems	Difference: Regular vs. Irregular
Stem Frequency			
B	0.056	0.029	0.027
t(131)	4.43	2.61	1.56
p	<.0001	.010	.120

Figure 1. Results of the linear regression model of stem acceptability ratings for regular and irregular verbs.

Past-Tense Acceptability

Analysis. Imageability, past-tense frequency, and the Frequency × Imageability interaction were entered into a linear regression model, with past-tense acceptability as the dependent variable. Stem acceptability ratings, number of phonemes in the past-tense forms, two measures of phonological neighborhood (i.e., of phonological similarity), and two lexical properties that have been shown to predict regularity (Baayen & Moscoso del Prado Martin, 2005) were included in the model as covariates; all six of these variables correlated with past-tense acceptability across the 137 regular and irregular verbs ($ps < .10$). Here and elsewhere, only variables that correlated with the dependent variable at the $p < .10$ level were included in the analyses. Each of the variables considered for inclusion in the model as a covariate is explained here.

Stem acceptability ratings was included as a covariate in order to account for the potential influence of stem access on past-tense acceptability, correlation with past-tense acceptability, $r(135) = .493$, $p < .0001$. Because the computation of a rule product is expected to involve stem access, any influence of the independent variables (imageability and past-tense frequency) on the stem of a rule product (e.g., *walk* in *walked*) could indirectly influence the inflected form itself and could potentially lead to a false conclusion that the inflected form is stored. Indeed, as we have seen above, imageability predicts stem acceptability. Moreover, past-tense frequency is highly correlated with stem frequency; for example, for the regulars and irregulars in this study, $r(135) = .925$, $p < .0001$, and this also predicts stem acceptability. Covarying out stem acceptability ratings thus ensures that any observed effects of the independent variables (imageability and past-tense frequency) on past-tense acceptability are in fact due to their influence on the past-tense forms rather than on their stems.

The number of phonemes in the past-tense forms was included as a covariate in order to account for the potential influence of the phonological length of these forms on their acceptability ratings. Word length has been shown to affect working memory performance (Caplan, Rochon, & Waters, 1992) and to predict performance on single-word processing measures such as lexical decision and single-word reading (Balota et al., 2004). The correlation between phoneme length and past-tense acceptability, $r(135) = -.143$, $p = .095$, indicated that shorter forms tended to be more acceptable.

Phonological neighborhood measures were included as covariates to account for the potential influence of phonologically similar verbs on a given verb stored in memory. On a connectionist perspective, the memory traces of both regular and irregular past tenses are affected by their "phonological neighborhood," which is a function of the type and token frequency of similar and dissimilar verbs. For example, the memory traces of *sing–sang* may be strengthened by *spring–sprang* and weakened by *bring–brought* and *wing–winged*. Some dual-system views also posit such neighborhood effects for forms stored in lexical memory (Pinker, 1999; Pinker & Ullman, 2002). We included two complementary neighborhood measures as covariates: "same-class" and "opposite-class" neighborhood strength, both of which correlated with past-tense acceptability: same-class neighborhood strength, $r(135) = .178$, $p = .038$; opposite-class neighborhood strength, $r(135) = .197$, $p = .021$.

Same-class neighborhood strength takes into account both "friends" and "enemies" in the same inflectional class (e.g., within irregulars). Same-class "friends" (which should strengthen memory traces) of a given irregular verb are defined as those irregulars that share the same type of stem–past transformation, specifically, those that are in the same "family," following Pinker and Prince (1988). For example, *feel–felt* and *leave–left* are both friends of *keep–kept*. Same-class "enemies" of a given irregular are defined as those irregulars that share the same rime with either the given verb or any of its same-class friends but have a different stem–past transformation and therefore are not in the same family (e.g., *steal–stole, weave–wove*). The same-class neighborhood strength of irregular verbs was calculated from the sum of the frequencies (FK + AP) of irregular friends minus the sum of the frequencies of irregular enemies. If this difference (which we will refer here to as D) was positive or zero, neighborhood strength was defined as $\ln(|D| + 1)$ (i.e., as the natural log of the absolute value of D); if negative, it was defined as $-\ln(|D| + 1)$, allowing a verb's neighborhood strength to be negative if it has a preponderance of irregular enemies. (The absolute value of D was augmented by 1 in each case to avoid ln(0).) The same-class neighborhood strength of regular verbs was calculated as the natural log of the sum of the frequencies of regular friends, again first augmented by 1 to avoid the undefined ln(0). Regular friends of a given regular are defined as those regular verbs that share the same vowel as the given verb. Note that regular verbs do not have regular enemies (i.e., enemies within the same inflectional class).

Opposite-class neighborhood strength was included to take into account potential regular neighborhood effects on irregulars and irregular neighborhood effects on regulars. Regular enemies of a given irregular verb were defined as those regular verbs with the same stem vowel as the irregular. Irregular enemies of a given regular verb were similarly defined as those irregular verbs with the same stem vowel as the regular. Note that although regulars that share their stem rime with any irregulars were excluded as inconsistent regulars (see above), we could not avoid at least some regulars that share the same stem vowel with at least some irregulars. For both irregulars and regulars, opposite-class neighborhood strength was calculated as the natural log of the sum of the frequencies of the enemies from the opposite inflectional class, first augmented by 1 to avoid ln(0).

Three lexical measures that have been shown to predict the regularity of English verbs (Baayen & Moscoso del Prado Martin, 2005) were also considered for inclusion as covariates in order to ensure that they do not explain any of the effects of interest. The first lexical measure that we considered is an estimate of the number of meanings of a given verb. This measure was calculated on the basis of information extracted from WordNet (Fellbaum, 1998; Miller, 1990), an online database of English words in which nouns, verbs, adjectives, and adverbs have been organized into synonym sets. Following Baayen and Moscoso del Prado Martin, we counted the number of verb synonym sets in which each of the regular and irregular verbs in this study appears. Because each synonym set represents one underlying lexical concept, this measure represents the number of different meanings associated with each verb. The measure correlated with past-tense acceptability, $r(135) = .278$, $p = .001$, and was therefore included in the model as a covariate.

The second lexically related measure is an estimate of the likelihood that a given verb has been converted from a noun or into a noun (Baayen & Moscoso del Prado Martin, 2005). Following Baayen and Moscoso del Prado Martin, we calculated the log of the ratio of a stem form's frequency as a noun to that form's frequency as a verb (based on the combined FK and AP frequency counts described above). This measure also correlated with past-tense acceptability, $r(135) = .290$, $p = .001$, and therefore was included in the regression model as a covariate.

The third lexical measure estimates the number of argument structures associated with a given verb. Following Baayen and Moscoso del Prado Martin (2005), this was calculated on the basis of Levin's (1993) classification of verbs into alternation classes, each of which represents a set of verbs that occurs with the same pattern of arguments and adjuncts. For each verb in this study, we calculated the natural log of the number of different alternation classes in which that verb appears (see Baayen & Moscoso del Prado Martin, 2005). However, this measure did not correlate with past-tense acceptability, $r(122) = -.009$, p = .918, and therefore was not included in the model as a covariate. (Note that the classification in Levin, 1993, did not include 13 of the verbs in Study 1—6 irregulars and 7 regulars—and therefore these verbs were not included in the correlation.) Other lexical variables explored in Baayen and Moscoso del Prado Martin were not available to us and therefore are not examined here.

The regression model that included these six covariates was designed to estimate the coefficients for past-tense frequency, imageability, and the Frequency × Imageability interaction, for both regular and irregular verbs. It yielded six coefficients in all (not including the six coefficients corresponding to the six covariates, which are not examined here) plus regular and irregular verb intercepts. The estimation of separate imageability, past-tense frequency, and Frequency × Imageability interaction coefficients for regulars and irregulars was accomplished by including in the model the interaction of each of these three terms with inflectional class. To avoid excessive complexity, we did not include interactions between inflectional class and the six covariates in the model. Both frequency and imageability were centered at their respective means; this allows the coefficient (main effect) of frequency to be interpreted as the estimated frequency effect at the mean imageability value and the coefficient (main effect) of imageability to be interpreted as the estimated imageability effect at the mean frequency value when the Frequency × Imageability interaction term was included in the model.

Regulars versus irregulars. Irregular past tenses showed significant main effects of both imageability and frequency (see Figure 2). In contrast, regulars showed neither effect. Regulars and irregulars differed significantly in the magnitude of both of their imageability and frequency effects. The Frequency × Imageability interaction term was included in the model because it was significant over both verb types at $p < .10$, $B = -0.024$, $t(123) = 1.92$, $p = .057$; note that the interaction did not differ between verb types, $B = -.010$, $t(123) = 0.40$, $p = .687$, so it was not examined separately for regulars and irregulars.

This pattern of frequency and imageability effects found for irregular but not regular past tenses does not seem to be explained

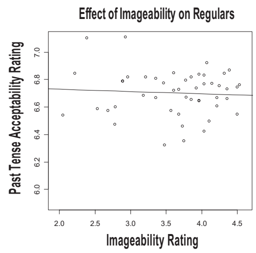
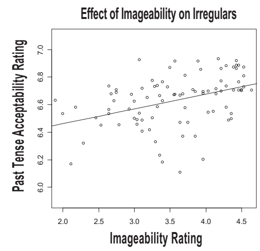
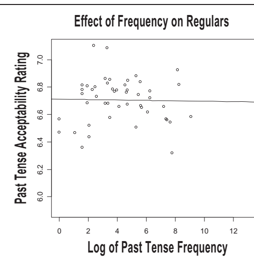
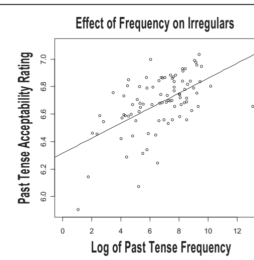
	Regular Past-Tenses	Irregular Past-Tenses	Difference: Regular vs. Irregular
Imageability			
B	-0.016	0.108	0.124
t(123)	0.34	3.12	2.12
p	.737	.002	.036
Past Tense Frequency			
B	-0.001	0.055	0.056
t(123)	0.11	4.47	3.49
p	.914	<.0001	.001

Figure 2. Results of the linear regression model of past-tense acceptability ratings for regular and irregular verbs.

by various experimental or statistical factors. First, it might be argued that imageability and frequency effects on the regulars may have been obscured by ceiling effects, due to high acceptability ratings for these forms. However, the past-tense acceptability ratings of regulars (range = 5.84–6.97, $M = 6.70$) and irregulars (range = 5.72–7.00, $M = 6.69$) did not differ significantly, $t(135) = 0.11$, $p = .91$; $F(88, 47) = 1.44$, $p = .17$, for the F test for equality of variance. Moreover, as we have seen above, imageability and frequency effects were found for the regulars' stem acceptability ratings, despite the fact that the regulars' mean stem acceptability rating (range = 5.56–7.0, $M = 6.72$) and past-tense acceptability rating were virtually identical. Finally, the fact that regular past-tense acceptability was in fact predicted by stem acceptability, $B = 0.403$, $t(125) = 3.57$, $p < .001$, shows that regular past-tense acceptability had sufficient variability for at least this predictor to reach significance in the regression model.[6] Thus it does not appear that past-tense acceptability ceiling effects explain the lack of imageability or frequency effects on the regulars.

Second, it might be suggested that the orthographic presence of the *-ed* affix among regular but not irregular past-tense forms might have encouraged more shallow processing of the regular forms. On this view, the lack of frequency and imageability effects on regular past-tense acceptability ratings is not due to the morphological parsing of regulars but to their more shallow processing in general, perhaps even at the orthographic level. However, the fact that regular past-tense acceptability ratings were predicted by stem acceptability (see just above) argues against this possibility, as such a relationship suggests that these past-tense forms were indeed processed deeply enough for their stems to be accessed.

Third, the inclusion of stem acceptability as a covariate might plausibly have led to a problem of collinearity between stem acceptability and either imageability or past-tense frequency. Because the coefficients of both imageability and stem frequency (which correlates highly with past-tense frequency; see above) were larger for regular than irregular stem acceptability ratings (see above), the inclusion of stem acceptability as a covariate could potentially weaken imageability and frequency effects more for regular than irregular past tenses if in fact stem acceptability was collinear with either imageability or past-tense frequency. However, stem acceptability was not collinear (i.e., $r > .8$; Berry &

[6] Note that irregular past tense acceptability ratings were also predicted by stem acceptability ratings, $B = 0.653$, $t(125) = 4.53$, $p < .0001$. Thus it is not the case that only regular past-tense acceptability is predicted by this measure of stem access. Whether or not the same and/or distinct explanations account for the stem acceptability effects observed for the two verb types may be clarified in future studies.

Felman, 1985) with either imageability or past-tense frequency for either regulars or irregulars ($rs < .47$).

Fourth, we considered the possibility of suppression (i.e., the phenomenon in which a change in the correlation coefficient between two independent variables influences the regression coefficients when both variables are included in a multiple regression model; Friedman & Wall, 2005). Because the correlation between imageability and frequency was different for regulars and irregulars (see below), suppression might explain the differences in frequency or imageability effects observed between the two verb types. However, when imageability was removed from the model, the pattern of frequency effects did not change: irregulars, $B = 0.054$, $t(127) = 4.34$, $p < .001$; regulars, $B = 0.001$, $t(127) = 0.08$, $p = .934$. Likewise, when frequency was removed from the model, the pattern of imageability effects was the same: irregulars, $B = 0.077$, $t(127) = 2.36$, $p = .020$; regulars, $B = 0.000$, $t(127) = 0.01$, $p = .995$. Thus the pattern of effects does not seem to be explained by suppression.

Fifth, it could be claimed that the smaller number of regular than irregular items, or the regulars' smaller past-tense frequency range (0.00–9.07 as compared to 1.10–13.07 for the irregulars) could lead to weaker effects for the regulars, due to insufficient power or variability. Additionally, the regulars had lower past-tense frequencies ($M = 4.21$) than did the irregulars ($M = 6.69$). We therefore compared the full set of regular verbs to the 40 lowest frequency irregulars, which were fewer in number than the 48 regulars, had a smaller past-tense frequency range (1.10–6.71), F test for equality of variance, $F(47, 39) = 2.60$, $p < .01$, and also did not differ significantly from the irregulars on past-tense frequency, $M = 4.81$, $t(86) = 1.44$, $p = .15$. As in the model with all 89 irregulars, the model with these 40 irregulars and the 48 regulars showed significant main effects of both imageability and frequency on irregulars but showed neither on regulars ($ps > .75$). The coefficients (imageability and frequency) differed significantly between the two verb types.[7] Thus the observed pattern does not seem to be explained by frequency or by set-size differences between regular and irregular verbs.

In sum, statistical and experimental factors do not seem to account for the pattern of imageability and frequency effects observed on irregular but not regular past-tense forms. The presence of imageability as well as frequency effects on irregulars is a new finding that confirms the utility of imageability effects (as well as frequency effects) as diagnostic of stored inflected forms. The absence of these effects among regulars is consistent with their rule-based composition and thus with a dual-system view. Although the observed pattern of findings has not, to our knowledge, been predicted or simulated by existing connectionist models, simulations from the domain of reading aloud (see Introduction) suggest the possibility that future connectionist models of inflectional morphology may be able to account for the findings.

Finally, we examined whether there might be a relationship between the imageability and past-tense frequency variables, independent of acceptability ratings or any other psycholinguistic measures. Dual-system theories might expect a specific relation between these two variables. On a dual-system view, irregular past tenses are preserved in a language diachronically (i.e., over the generations) only if they are sufficiently well remembered. Otherwise, an irregular verb may eventually become regular (e.g., *gilded*, as compared to *gilt*). This diachronic pattern can explain why the token frequencies of irregular past tenses (i.e., how often one encounters each irregular past-tense form) are much higher than those of regular past tenses (Pinker, 1999): Rarely used irregulars will tend to become regulars and higher frequency irregulars will be left in the language. If indeed imageability also plays a role in the memorization of irregulars, one might expect that lower frequency irregulars are more imageable and that less imageable irregulars have higher frequencies. In other words, even if an irregular past tense is of low frequency, it may be remembered if it is imageable enough, and a low-imageability irregular should be remembered if it is frequent enough. Thus one might expect a negative correlation between frequency and imageability in irregulars. Indeed, among the 89 irregular verbs in this study, past-tense frequency and imageability showed a negative correlation, $r(87) = -.245$, $p = .021$. In contrast, no such correlation should be expected for regulars on a dual-system view, and indeed none was found among the 48 regulars, $r(46) = .078$, $p = .597$. It might be argued that the absence of a correlation among the regulars may be due to the smaller sample size of regulars than irregulars. However, the analogous correlation over all 142 consistent regulars for which imageability ratings were acquired (out of 436 verbs; see above) showed the same pattern, $r(140) = -.073$ $p = .389$. This finding of a negative correlation between past-tense frequency and imageability for irregulars but not regulars, which constitutes an interesting source of additional evidence for dual-system models, suggests that imageability does indeed influence the diachronic retention of irregularly inflected forms.

Study 2

In Study 2 we explored imageability and frequency effects on response times in a past-tense production task.

Method

Past-tense production response times (RTs) were acquired from 72 subjects (36 males and 36 females). All subjects were right-handed native speakers of English, with no personal history of psychiatric, neurological, or learning disorders. The subjects ranged in age from 18 to 50 years ($M = 25$ years, $SD = 8$) and had a minimum of 12 years of education (maximum = 21; $M = 15$ years, $SD = 2$). One subject's mean RT was greater than two standard deviations from the mean calculated over all nonexcluded real words (see below for excluded items) and was thus eliminated from all analyses.

Each subject was tested on 112 verbs: 32 irregulars, 32 consistent regulars, 16 inconsistent regulars, and 32 novel verbs. Only

[7] This set of 40 irregulars was obtained by first selecting those 45 irregulars below the irregular past-tense frequency median and then eliminating the 10% highest frequency remaining verbs, as the initial set of 45 verbs below the median still had significantly higher past-tense frequencies than did the regulars. Note that in this model (i.e., with 40 irregulars and 48 regulars) there was no interaction between imageability and frequency over both verb types and no difference in the interaction term between the verb types ($ps > .64$); therefore, the interaction term was not included in the final model. Finally, note that although the interaction between imageability and frequency over both verb types was borderline significant in the original model ($p = .057$), its lack of significance in this model suggests the absence of a real effect.

the irregular and consistent regular verbs are analyzed or discussed here (for a full list, see Walenski, Mostofsky, and Ullman, 2007). The stems of the consistent regulars shared their rimes with the stems of few or no irregulars. The irregulars did not include any no-change verbs (e.g., *hit–hit*) or doublet verbs (e.g., *dive–dived/dove*). These 32 regulars and 32 irregulars were matched pairwise on stem (unmarked form) frequency, regulars, range = 1.10–8.58, $M = 5.70$; irregulars, range = 1.79–9.02, $M = 5.67$; $t(31) = 0.08$, $p = .93$, and on past-tense frequency, regulars, range = 1.61–9.43, $M = 5.86$; irregulars, range = 1.79–10.92, $M = 5.73$; $t(31) = 0.77$, $p = .45$. The regular and irregular verbs also did not differ on imageability: regulars, range = 1.96–4.56, $M = 3.45$; irregulars, range = 2.18–4.64, $M = 3.57$; $t(62) = 0.65$, $p = .52$.

Three regular/irregular pairs were excluded from analyses: *Sink* is a doublet verb, with two possible irregular past tenses (*sank, sunk*); the past tense of *bind* is also a verb stem (*bound*). *Owe* is the only single-phoneme verb and moreover shares its rime with many more irregular stems than does any other verb selected as a consistent regular (*owe* rhymes with seven irregular stems; all other consistent regulars in this study rhyme with between zero and three irregulars). These three verbs and the pairwise matched verbs from the opposite inflectional class (*sign, drown*, and *stride*, respectively) were removed from the data set prior to all analyses.[8] Like the full set of 32 pairs, the remaining 29 pairs of regulars and irregulars did not differ statistically on stem frequency, past-tense frequency, or imageability ($ps > .40$).

Each verb stem was presented alone and in the context of a sentence, with a second sentence eliciting the past-tense form (e.g., *fail. Every day I fail an exam. Just like every day, yesterday I ____ an exam*). Other than in the stimulus verb (e.g., *fail*) and the words in the sentence that followed it (e.g., *an exam*), the sentences did not differ. All postverbal complements or adjuncts were composed of two words, neither of which was inflected or of low frequency. The verb stem and the two sentences were displayed at the same time on the computer screen, one below the other. Subjects were instructed to produce the missing form as quickly and accurately as possible. Items were pseudorandomized, and all subjects were presented with the items in the same order (see below).

Analyses were performed only on correct first responses (96.4% of all first responses to the 64 regular and irregular verbs). The experimenter monitored the microphone trigger and noted items whose RTs were not triggered by the subject's response; these items were discarded (1% of RTs for real verbs). RTs faster than 500 ms were discarded as being likely due to computer error (2.1% of correct responses). Extreme outliers for each subject (i.e., responses whose RTs were more than 3.5 standard deviations from the given subject's real-word mean) were removed for that subject (1.0% of correct responses).

Analysis

The RT data were analyzed using hierarchical linear modeling (HLM) with subject as a random factor. This statistical method allows each individual RT from each subject to be entered into one model without the RTs having been averaged across subjects for each item; this averaging results in a substantial loss of information. Additionally, HLM allows one to account for subject variability by estimating a separate intercept for each subject. For more complete discussions on this statistical method for similar analyses, see Baayen (2004) and Baayen, Davidson, and Bates (2008).

Imageability, frequency, and the Frequency × Imageability interaction were entered as item-level (Level 1) variables in a hierarchical linear model, with natural log–transformed RT as the dependent variable. Five item-level (Level 1) covariates were included in the model. As in Study 1, only covariates that correlated with the dependent variable (here, log of RT) at $p < .10$ were included in the model. The five covariates were item order; number of phonemes in the verb stem; whether or not the onset of the verb was a fricative; whether or not the previous verb was of the same inflectional class (i.e., regular or irregular); and whether the previous verb was real or novel. Each of these covariates, as well as several variables that were considered but were not included in the model, is explained here.

First, we included a measure of item order (i.e., of the number of verbs already presented before a given verb). Explicitly including item order in the model allows one to directly account for variability attributable to presentation order (e.g., due to practice effects within the task), which in turn enhances prediction accuracy and decreases the size of the residual error. Including order as a covariate allows for a single presentation order; this is advantageous with respect to simplicity of task administration and response coding. Order is likely to be most influential for the first few items, as order effects diminish as the subjects become more comfortable with the task. Therefore, rather than covary out item order itself, we covaried out the natural log of item order, which indeed correlated significantly with log RT, $r(62) = -.383$, $p = .002$.

The phonological length of each item is also likely to affect RT, at the very least because phonological length affects working memory (Baddeley & Hitch, 1974; Caplan et al., 1992), and the stem is likely to be held in working memory before production of the past-tense form. Indeed, stem phoneme length (the number of phonemes in the presented stem) correlated with log RT, $r(62) = .272$, $p = .030$, but past-tense phoneme length did not, $r(62) = .141$, $p = .265$. Therefore, stem phoneme length was included as a covariate.

Three binary variables were included as covariates: (a) whether or not the onset of the verb was a fricative, correlation with log RT, $r(62) = .284$, $p = .023$, because this contrast can affect the timing of microphone triggering (Kessler, Treiman, & Mullennix, 2002); (b) whether or not the previously presented verb was of the same inflectional class (i.e., regular or irregular), correlation with log RT, $r(62) = -.264$, $p = .035$, because repeating a similar response or producing a different type of response may affect RT; and (c) whether the previous verb was real or novel, correlation with log RT, $r(62) = -.405$, $p < .001$, because switching from a novel to a real response could also affect processing time.

Finally, the two measures of phonological neighborhood (same-class neighborhood strength and opposite-class neighborhood

[8] The 3 irregular items excluded in Study 2 (*sink, bind,* and *stride*) were included in analyses in Study 1. (The 3 regulars excluded in Study 2—*sign, drown,* and *owe*—were not in the set of 48 consistent regulars in Study 1.) We did not exclude these items in Study 1, because the number of irregulars in that study was much larger than in Study 2 (89 vs. 32) and thus noise was less likely to affect the results. Nevertheless, we examined the regression model excluding these 3 items (i.e., on 86 irregulars and 48 regulars); this model yielded the same pattern of results.

strength) and the three lexical properties that predict regularity (number of synonym sets to which a given verb belongs, noun-to-verb frequency ratio, and the number of alternation classes to which a verb belongs) that were examined for inclusion as covariates in Study 1 were also considered here. However, because none of these variables correlated with log RT ($ps > .12$), none were included in the model.

As in Study 1, both regular and irregular verbs were included in all models, and separate coefficients for frequency and imageability were generated for each inflectional class (as were an intercept for each inflectional class and a single coefficient for each covariate, not reported here). Frequency and imageability were always centered on the mean of all 58 verbs.

Although the inclusion of variables in each model was based on various considerations other than model fit (e.g., for covariates, correlations with the dependent variable), the specification of random effects was determined by model fit. In the model including all 58 regular and irregular verbs, adding a random effect of subject on the intercept dramatically improved model fit, from Bayesian information criterion (BIC) = 4,366.6 to BIC = 229.6 (smaller is better; for an explanation of BIC, see Hox, 2002). Of the five covariates and three variables of interest (imageability, frequency, and verb type), only log of order considerably improved model fit (from BIC = 229.6 to BIC = 116.3) when its coefficient was allowed to vary by subject. Therefore, the final model included random effects of subject both on intercept and on the coefficient of log of order, with all other effects fixed (constant coefficients). We retained the same random effects and covariate structure in all models in Study 2. Note that no explicit subject-level (Level 2) variables (i.e., other than subject itself) were initially included; in most analyses, the subject level serves only to model between-subject variation, both in baseline RT and in the effects of log of order. Only when the subject-level variable sex (male vs. female) is added to the model (see analyses below) is an explicit Level 2 variable taken into account.

Regulars Versus Irregulars

The irregulars yielded significant negative coefficients of imageability and past-tense frequency. That is, higher frequency and higher imageability irregulars were produced faster (i.e., with smaller RTs) than lower frequency and lower imageability irregulars (see Figure 3). In contrast, the regulars showed no imageability effect. Moreover, as expected, the magnitude of the imageability effect on the regulars differed significantly from that on the irregulars. However, unlike in Study 1, frequency yielded a significant negative coefficient for regulars that indicated frequency effects for these forms (see below for discussion and further analysis). Nevertheless, as expected, the frequency effect for irregulars was significantly stronger than that for regulars (see Figure 3). Note that the Frequency × Imageability interaction was included in the model, as it was significant over both verb types, $B = -0.008$, $t(3732) = 2.66$, $p = .008$;[9] however, the interaction did not differ significantly between the verb types, $B = -0.008$, $t(3729) = 1.47$, $p = .140$, and thus was not examined separately for regulars and irregulars.

As in Study 1, the imageability and frequency effect differences between regulars and irregulars do not seem to be explained by experimental or statistical factors. The mean RT for regulars did not differ from the mean RT for irregulars, which suggests that the weaker effects on regulars are not due to ceiling effects: regulars, range = 1,447–2,328, $M = 1,772$; irregulars, range = 1,479–2,172, $M = 1,752$; $t(56) = 0.36$, $p = .72$; $F(28, 28) = 1.26$, $p = .55$, for the F test for equality of variance. Additionally, as we have seen above, the two verb types were matched on stem frequency, past-tense frequency, and imageability; moreover, multiple potential explanatory variables were covaried out. Thus the finding of imageability and frequency effects on irregulars confirms the results from Study 1, as does the lack of imageability effects on regulars.

However, the presence of frequency effects on regulars suggests at least some storage of these forms. Here we examine whether this pattern may be at least partly explained by the inclusion of female subjects (see Introduction and Ullman, 2004; Ullman et al., 2008).[10] We therefore added the subject-level (Level 2) factor sex (male vs. female) to the model and included the Sex × Verb Type × Imageability and the Sex × Verb Type × Frequency interaction terms in order to generate separate imageability and frequency coefficients for males and females on regular and irregular items. We included the same covariates as in the model reported above. To avoid excessive complexity in the model, we added no interaction terms between the covariates and either sex or verb type.

The interaction between sex, verb type, and frequency was significant, $B = 0.016$, $t(3750) = 2.15$, $p = .031$, which indicated that frequency effects differed between the verb types in the two sexes (i.e., frequency effects differed among female irregulars, male irregulars, female regulars, and male regulars). The interaction between sex, verb type, and imageability was also significant, $B = 0.043$, $t(3721) = 2.00$, $p = .045$, indicating that imageability effects also differed between regulars and irregulars in the two sexes.x[11] Note that the Frequency × Imageability interaction over both verb types and sexes was significant, $B = -0.008$, $t(3725) = 2.65$, $p = .008$, and thus the interaction term was included in the model. However, the four-way interaction between sex, verb type, frequency, and imageability was not significant, $B = -0.003$, $t(3722) = 0.25$, $p = .800$, so Frequency × Imageability interaction terms were not examined separately by sex or verb type.

---

[9] The significant Frequency × Imageability interaction found over both verb types showed that as past-tense frequency increased, the imageability effect became stronger and that as imageability increased, the frequency effect became stronger. This finding contrasts with Study 1, in which no reliable Frequency × Imageability interaction was observed. Future studies may clarify the nature of this difference.

[10] Note that in Study 1 we did not explore frequency and imageability effects separately in male and female subjects because we did not have subject information, including sex, for this task (see Footnote 2).

[11] Tabak, Schreuder, and Baayen (2005) found that sex differences in inflectional morphology disappeared when the frequency coefficient was allowed to vary by subject (i.e., adding by-subject random slopes for frequency to the model). However, in the present study, the inclusion of by-subject random slopes for either frequency or imageability did not improve model fit (in fact, model fit worsened slightly; from BIC = 160.1 to 169.4 and to 170.6, for frequency and imageability, respectively), and the pattern of significance produced by the models did not change at all. Most important, and unlike in Tabak et al., in the model with by-subject random slopes for frequency, the Frequency × Sex × Verb Type interaction remained significant, $B = 0.017$, $t(3703) = 2.19$, $p = .029$, and in the model with by-subject random slopes for imageability, the Imageability × Sex × Verb Type interaction remained significant, $B = 0.042$, $t(3665) = 1.99$, $p = .047$.

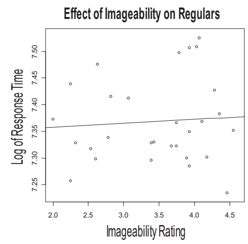
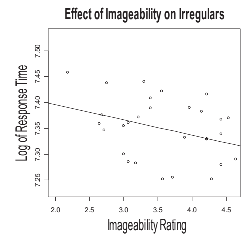
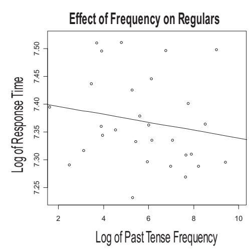
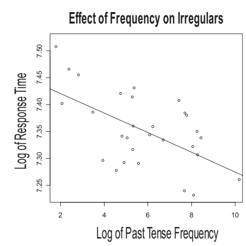
	Regular Past-Tenses	Irregular Past-Tenses	Difference: Regular vs. Irregular
			
Imageability			
B	0.007	-0.029	0.037
t	*t*(3728) = 1.03	*t*(3729) = 3.54	*t*(3728) = 3.41
p	.303	.0004	.0007
			
Past Tense Frequency			
B	-0.007	-0.018	0.011
t	*t*(3729) = 2.49	*t*(3730) = 6.26	*t*(3730) = 2.80
p	.013	<.0001	.005

Figure 3. Results of the hierarchical linear model of past-tense production response times for regular and irregular verbs.

As can be seen in Figure 4, the males showed both frequency and imageability effects on irregulars but not on regulars,[12] and the magnitude of both effects differed significantly between the verb types. In contrast, the females showed no differences between regulars and irregulars in the magnitude of either frequency or imageability effects, with significant frequency effects for both verb types and no significant imageability effects for either. In addition, on irregular verbs, males and females did not differ on either their imageability coefficients, $B = 0.023$, $t(3767) = 1.39$, $p = .166$, or their frequency coeffiecients, $B = 0.004$, $t(3763) = 0.64$, $p = .521$. However, females showed significantly stronger frequency effects for regulars than did males, $B = 0.013$, $t(3766) = 2.37$, $p = .018$, though this sex difference was not significant for the imageability coefficients, $B = 0.020$, $t(3779) = 1.42$, $p = .155$.

The male/female differences observed here do not seem to be explained by factors unrelated to sex. The two sexes did not differ on handedness (all right-handed); age (females, range = 18–50, $M = 26.06$ years; males, range = 18–50, $M = 23.92$ years), $t(69) = 1.15$, $p = .25$; or education (females, range = 12–20, $M = 15.66$ years; males, range = 12–21, $M = 15.03$ years), $t(69) = 1.49$, $p = .14$. As stated above, all subjects had similar linguistic histories, and none of them had any psychiatric, neurological, or learning disorders. There were also no sex differences in overall RT over regular and irregular items: males, range = 1,601–2,022, $M = 1{,}811$; females, range = 1,490–1,932, $M = 1{,}711$, $t(70) = 0.67$, $p = .506$; this argues against the possibility that one sex was simply faster than the other (e.g., due to a greater interest in the task). Moreover, note that any potentially confounding subject-related difference between males and females would need to

[12] The slope for imageability on regulars in males was actually positive, with the effect approaching significance ($p = .085$; see Figure 4). This suggests the possibility of "anti-imageability" effects. Antifrequency effects have previously been reported for regulars but not irregulars in several studies (Alegre & Gordon, 1999; Prasada et al., 1990; Ullman, 1993, 1999). We have suggested that anti-frequency effects may reflect compositional processes: Even if inflected forms are composed, previous encounters with these forms are likely to leave some representation in lexical memory. The higher the frequency of any such forms, the stronger their memory traces. Even though the memory traces may be too weak to result in the successful retrieval of the form, stronger memory traces should lead to a greater attempt at retrieving the form and increase its processing time. Therefore, paradoxically, for forms that are ultimately computed by the rule system, those with higher frequencies may take longer for the lexical system to reject and so will inhibit the rule system for a longer period of time. Anti-imageability effects, if proven reliable in future studies with more power (e.g., with more items), may be explained in an analogous manner.

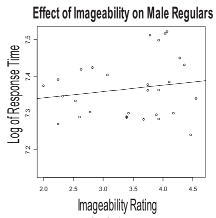
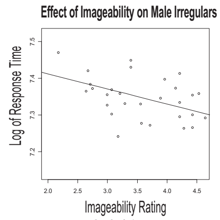
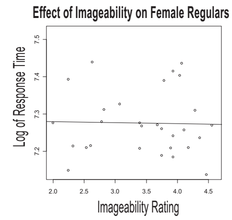
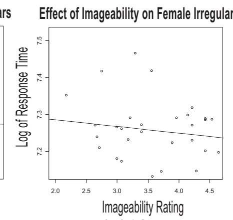
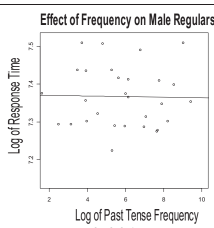
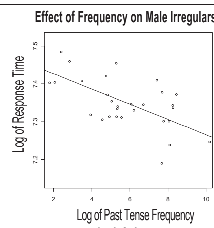
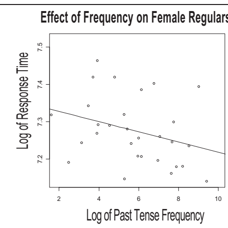
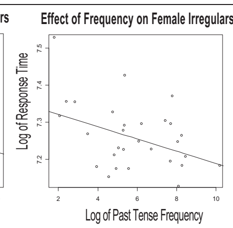
	Male			Female		
	Regular	Irregular	Difference: Regular vs. Irregular	Regular	Irregular	Difference: Regular vs. Irregular
Imageability						
B	0.017	-0.041	0.058	-0.003	-0.018	0.015
t	*t*(3757) = 1.73	*t*(3749) = 3.50	*t*(3722) = 3.84	t(3754) = 0.26	*t*(3745) = 1.55	*t*(3721) = 1.01
p	.085	.001	.0001	.793	.121	.311
Past Tense Frequency						
B	0.001	-0.020	0.019	-0.014	-0.017	0.003
t	*t*(3747) = 0.20	*t*(3787) = 5.00	*t*(3738) = 3.53	t(3744) = 3.44	*t*(3788) = 4.07	*t*(3737) = 0.52
p	.840	<.0001	.0004	.001	<.0001	.600

Figure 4. Results of the hierarchical linear model of past-tense production response times for males and females for regular and irregular verbs.

explain the specific pattern that was observed (i.e., Sex × Verb Type interactions for both frequency and imageability effects).

In sum, Sex × Verb Type interactions were obtained for both frequency and imageability effects. Males but not females showed more reliable frequency and imageability effects for irregulars than regulars. Overall, this pattern, which is not explained by a variety of potentially confounding factors, is consistent with the hypothesis that males and females process inflected forms differently and, more specifically, that males but not females are more likely to store irregular than regular inflected forms (Hartshorne & Ullman, 2006; Ullman, 2004, 2005a; Ullman et al., 2008).

However, at low enough frequencies, even females may be expected to have difficulty memorizing regularly inflected forms, given that they have had little exposure to them (see Introduction). We therefore examined this hypothesis (i.e., whether females may be more likely to store and retrieve high-frequency than low-frequency regular inflected forms as well as high-imageability as compared to low-imageability regulars).

To test this hypothesis, we added quadratic frequency and imageability terms to the model. If females retrieve only high-frequency and/or high-imageability regulars from memory, one might expect the quadratic terms to be significant and negative and thus to produce a concave-downwards curve, consistent with frequency effects primarily for higher frequency items. In contrast, no such pattern would be expected for irregulars.

Indeed, the Frequency × Frequency × Verb Type interaction in females was significant ($p = .017$), with regulars but not irregulars yielding a significant negative quadratic term that indicated a concave-downwards curve (see Figure 5). The Imageability × Imageability × Verb Type interaction approached significance ($p = .073$), with a negative quadratic term for regulars also approaching significance (see Figure 5).[13]

The significant negative frequency quadratic term for regulars could be explained by the regression line curving downward in the high-frequency range, indicating frequency effects on higher frequency regulars, and/or by the regression line curving upward in the lower frequency range, indicating anti-frequency effects among low-frequency regulars (see Footnote 12). Visual inspection of the curve suggests both of these possibilities. We therefore tested whether the slope was significantly different from zero at both the highest frequency item (log frequency = 9.43) and the lowest frequency item (log frequency = 1.61). Note that the linear frequency terms reported in Figure 5 represent the linear frequency effects at the mean frequency value, because, as stated above, frequency was centered on its mean. Note also that we examined the slope at the highest and lowest frequency points, rather than points outside the data set, to avoid extrapolating beyond the data. At the highest frequency item, the slope was negative and significantly different from zero, indicating significant frequency effects, $B = -0.040$, $t(3740) = 2.83$, $p = .005$. In contrast, although the slope at the lowest frequency item was positive, it did not differ significantly from zero: frequency, $B = 0.016$, $t(3719) = 1.05$, $p = .296$. These results support the hypothesis that females are more likely to retrieve high- than low-frequency regulars. Note that the analogous analysis on the imageability quadratic for regulars yielded a similar pattern, although at the highest imageability item

[13] None of the frequency or imageability quadratic terms or their interactions with verb type were significant for males.

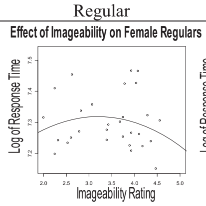
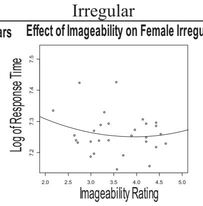
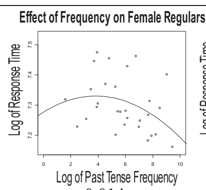
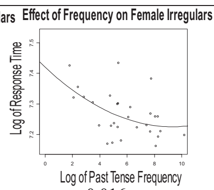
	Regular	Irregular	Difference: Regular vs. Irregular
Imageability			
B (linear)	-0.019	-0.013	0.006
t (linear)	*t*(3737) = 1.44	*t*(3763) = 1.04	*t*(3718) = 0.31
p (linear)	.149	.296	.755
B (quadratic)	-0.029	0.016	0.045
t (quadratic)	*t*(3718) = 1.72	*t*(3741) = 0.85	*t*(3740) = 1.79
p (quadratic)	.086	.398	.073
Past Tense Frequency			
B (linear)	-0.014	-0.016	-0.001
t (linear)	*t*(3744) = 3.58	*t*(3780) = 3.78	*t*(3737) = 0.25
p (linear)	<.001	<.001	.802
B (quadratic)	-0.003	0.002	0.006
t (quadratic)	*t*(3729) = 1.98	*t*(3781) = 1.34	*t*(3731) = 2.40
p (quadratic)	.048	.180	.017

Figure 5. Results of the hierarchical linear model of past-tense production response times among females for regular and irregular verbs, with quadratic frequency and imageability terms included.

the difference between the negative slope and zero only approached significance: highest imageability item, $B = -0.080$, $t(3714) = 1.80$, $p = .072$; lowest imageability item, $B = 0.071$, $t(3726) = 1.55$, $p = .121$. Future studies with additional power should reveal whether or not the near-significant effects related to the imageability quadratic on regulars (here and above) are reliable.

In sum, the finding that females (but not males) showed significant frequency effects on regulars, particularly among higher frequency items, at least partially explains the finding that frequency effects were observed on regulars in the analyses performed over both sexes (see Figure 3). This pattern of significance on regulars was reliable only for frequency, not for imageability. That is, over all subjects (see Figure 3) and in females (see Figures 4 and 5), regulars showed significant frequency but not imageability effects. One possibility is that, in general, frequency contributes more than does imageability to the storage of inflected—and possibly other linguistic—forms. The finding that in females, irregulars as well as regulars yielded a pattern of significant frequency but not imageability effects (see Figure 4) seems to be consistent with this view. However, it may instead or additionally be the case that frequency has a greater influence than imageability in females but not (or less so) in males. This might result, for example, from a female advantage in "verbal" processing and a male advantage at "visuospatial" processing (Kimura, 1996, 1999; Lewin, Wolgers, & Herlitz, 2001; Sherwin, 2003), which could lead to a particular female reliance on the frequency of linguistic forms and a relatively larger male reliance on the imageability of these forms. Indeed, the finding that, unlike females, males showed robust imageability as well as frequency effects on irregulars is consistent with this hypothesis, which can be further elucidated in future studies.

Summary and Discussion

The results from the two studies can be summarized as follows. First of all, both regular and irregular verb stems showed both frequency and imageability effects (Study 1), validating imageability effects alongside frequency effects as a diagnostic measure of the storage of linguistic forms. More important, in Studies 1 and 2, irregular inflected forms showed both imageability and frequency effects, which were moreover significantly stronger than the imageability and frequency coefficients for regulars. In contrast, regular inflected forms in both studies showed either no imageability or frequency effects at all (Study 1) or only frequency effects (Study 2).

We next examined the hypothesis that the frequency effect on regulars in Study 2 may have been partly explained by the inclusion of female subjects (see Introduction). The addition of sex (male vs. female) as a factor in the model elicited Sex $\times$ Verb Type interactions for both frequency and imageability; males but not females showed more reliable frequency and imageability effects for irregulars than regulars, and females but not males showed frequency effects on regulars (see Figure 4). Finally, to examine whether the frequency effect on regulars in females was due primarily to the presence of higher frequency forms (see Introduction), we introduced quadratic terms, which revealed that the frequency effect on regulars in females held only for higher frequency verbs (see Figure 5).

The data obtained in these studies constitute various novel findings, including the imageability effects, sex differences, and quadratic frequency effects. These findings are informative and potentially important independent of any theoretical framework. In addition, they appear to restrict the theoretical landscape.

On the one hand, as discussed above, some of the results are consistent with more than one theoretical perspective. For example, the imageability and frequency effects observed on irregulars seem to be broadly consistent with single-mechanism and dual-system models, both of which assume the storage of some sort of representations for irregulars. However, it remains to be seen how some hypotheses (e.g., Halle & Marantz, 1993) might account for the imageability effects. Moreover, the observation that imageability effects were reliable on irregulars in males but not females seems inconsistent with the hypothesis that females might be expected to show stronger imageability effects than do males on a single-mechanism perspective (see Introduction).

Additionally, the pattern of effects on regulars, in particular the presence of frequency effects on regulars in females but not males and the significant quadratic frequency term (as well as the near-significant quadratic imageability term) on regulars among females, seems to be consistent with only certain theoretical perspectives. First, these findings seem to be incompatible with dual-system claims that no regulars (Marslen-Wilson & Tyler, 1998; Taft, 1979; Tyler et al., 2002), or all previously encountered regulars (Caramazza et al., 1988), are stored. Second, this pattern of results does not seem to be predicted by single-mechanism perspectives (Bybee, 1995; Joanisse & Seidenberg, 1999; McClelland & Patterson, 2002). No single-mechanism model that we are aware of has predicted frequency effects on regulars among females but not males or frequency effects among high- but not low-frequency items. Indeed, as we discuss in the Introduction, such models might be expected to yield a different pattern, such as weaker effects in females than males. It remains to be seen whether future connectionist models can accommodate the data presented here.

In contrast, the pattern of results observed here is expected by certain dual-system views, specifically, those models in which regularity, frequency, imageability, and sex influence the likelihood of storage of inflected forms (Alegre & Gordon, 1999; Baayen et al., 1997; Hartshorne & Ullman, 2006; Pinker, 1999; Pinker & Ullman, 2002; Ullman, 1993, 2001a, 2004, 2005a; Ullman, Walenski, Prado, Ozawa, & Steinhauer, under revision; Walenski, Mostofsky, Larson, & Ullman, 2008). Certain results obtained here have been predicted only, to our knowledge, by the dual-system perspective that lexical items are stored in the "declarative memory" system, which also underlies semantic knowledge. In particular, this view predicts that the observed sex differences, which follow from an apparent female advantage at aspects of declarative memory, lead to a greater likelihood that females will store regulars and other complex forms (Hartshorne & Ullman, 2006; Ullman, Miranda, & Travers, 2008; Ullman, 2004, 2005a; Ullman et al., 2008). Note, however, that the sex differences reported here may be compatible with other dual-system models, although it remains for such models to address this issue.

The pattern of frequency and imageability effects found across the two studies is consistent with previously reported findings of frequency effects being more reliable on irregular than regular inflected forms (Prasada, Pinker, & Snyder., 1990; Seidenberg & Bruck, 1990; Ullman, 1999; van der Lely & Ullman, 2001) and on higher than lower frequency regulars (Alegre & Gordon, 1999; unlike in Study 2, however, Alegre and Gordon used an arbitrary cutoff between high and low frequency forms). Ours is the first experiment, to our knowledge, that has reported more reliable frequency effects in females than in males, although other psycholinguistic evidence also supports the more reliable storage of regulars in females (Hartshorne & Ullman, 2006; Ullman et al., 2008; Ullman, Walenski, et al., under revision). Previous reports of frequency effects on regulars may be explained by the inclusion of female subjects or of regular items that are inconsistent or of high frequency (Marchman, 1997; Sereno & Jongman, 1997; Stemberger & MacWhinney, 1988; for different findings, in Dutch, see Tabak, Schreuder, & Baayen, 2005).

Together, the data from previously reported studies and those presented here suggest that frequency, imageability, and sex influence the storage of inflected forms. These factors seem to facilitate the memorization not only of irregulars and other idiosyncratic forms (e.g., verb stems), which must be stored, but of regulars and possibly other complex forms, which may be stored but can also be composed according to rule-governed operations. Thus, at least English regular past-tense forms can apparently be represented in more than one way. This finding underscores the importance of redundant representations in language (Ullman, 2005b, 2007; Walenski & Ullman, 2005).

More generally, the evidence suggests that the memorization of both simple and complex forms depends on the interaction of various factors, including lexical properties of the items being processed and characteristics of individual speakers. Careful examination of the interplay of these factors seems critical for the advancement of our understanding of the representation and processing of both simple and complex linguistic forms. Thus we believe that future studies should control for, and ideally directly examine the influence of, all factors that may be likely to play a role in the storage of linguistic forms.

A number of other issues remain to be resolved. First, the precise nature of the effect of imageability on storage is not yet well understood, even for the uninflected forms that have been investigated in previous studies of imageability. For example, it is not yet clear whether imageability facilitates initial memorization and/or later access or retrieval of forms. The psychological mechanisms by which imageability might facilitate storage or recall (e.g., mental imagery, context availability, semantic richness) are also still under debate (Binder, Westbury, McKiernan, Possing, & Medler, 2005; Jones, 1985; Kieras, 1978; Paivio, 1971, 1995; Sabsevitz, Medler, Seidenberg, & Binder, 2005; Schwanenflugel, Akin, & Luh, 1992; Schwanenflugel & Shoben, 1983), as is the relation between imageability and related factors, such as concreteness (Paivio, Yuille, & Madigan, 1968) or "embodiment" (Baayen, 2007).

Similarly, the processes involved in the observed effects of sex and even frequency are still not fully understood. For example, it is not yet clear to what extent the apparent female advantage at remembering lexical forms is one of memorization and/or recall and whether this advantage is due to sex differences in hormonal status (e.g., estrogen), physiology (e.g., dendritic branching), anatomy (e.g., hippocampal volume), environmental variables (e.g., exposure to language), or some combination of these and other factors (Miranda & Ullman, under review; Ullman, 2004, 2005a; Ullman et al., 2008). Additionally, the precise relation between frequency and age of acquisition on the storage of linguistic forms is still not resolved (Brown & Watson, 1987; Gerhand & Barry, 1998, 1999; Morrison & Ellis, 1995).

The studies reported here suggest new directions for research. For example, the findings suggest the possibility that imageability may facilitate the storage not only of inflected forms but also of phrases or even sentences. Indeed, some evidence supports this hypothesis. In one study, concrete sentences were comprehended faster and recalled more accurately than abstract sentences (Holmes & Langford, 1976). In another, subjects were better at identifying both wording and meaning changes in concrete than in abstract sentences (Moeser, 1974; but see Begg & Paivio, 1969). It remains to future research to further explore such imageability effects on complex syntactic structures and their potential interactions with other factors.

In conclusion, we have introduced imageability effects as a new diagnostic measure of the storage of linguistic forms and have provided converging evidence from both imageability and frequency effects, across different dependent measures. It is argued that this evidence supports a dual-system view in which the line between storage and composition is not at all static. Rather, it is highly dynamic, depending upon the interplay of both item- and subject-level factors, including frequency, imageability, and sex, all of which may affect the storage of simple and complex forms.

References

Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language, 40,* 41–61.

Baayen, H. R. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers, 1,* 1–47.

Baayen, R. H. (2007). Storage and computation in the mental lexicon. In G. Jarema & G. Libben (Eds.), *The mental lexicon: Core perspectives* (pp. 81–104). Amsterdam: Elsevier.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language: Emerging Data Analysis Techniques, 58,* 390–412.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language, 37,* 94–117.

Baayen, R. H., & Moscoso del Prado Martin, F. (2005). Semantic density and past-tense formation in three Germanic languages. *Language, 81,* 666–698.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283–316.

Begg, I., & Paivio, A. (1969). Concreteness and imagery in sentence meaning. *Journal of Verbal Learning and Verbal Behavior, 8,* 821–827.

Berry, W. D., & Felman, S. (1985). *Multiple regression in practice.* Beverly Hills, CA: Sage.

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuropsychology, 17,* 905–917.

Brown, G. D., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory and Cognition, 15,* 208–216.

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

Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *Journal of Experimental Psychology, 45,* 177–192.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28,* 297–332.

Chiarello, C., Shears, C., & Lund, K. (1999). Imageability and distributional typicality measures of nouns and verbs in contemporary English. *Behavior Research Methods, Instruments, & Computers, 31,* 603–637.

Church, K. (1988, February). *A stochastic parts program and noun phrase parser for unrestricted text.* Paper presented at the Second Conference on Applied Natural Language Processing, Austin, TX.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences, 22,* 991–1060.

Daugherty, K., & Seidenberg, M. (1992, April). *Rules or connections? The past tense revisited.* Paper presented at the Milwaukee Rules Conference, Milwaukee, WI.

Elman, J. L. (1996). Rethinking innateness. In J. L. Elman (Ed.), *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Embick, D., & Marantz, A. (2005). Cognitive neuroscience and the English past tense; comments on the paper by Ullman et al. *Brain and Language, 93,* 243–247.

Fellbaum, C. E. (1998). *WordNet: An electronic database.* Cambridge, MA: MIT Press.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12,* 627–635.

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

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 361–379.

Friedman, L., & Wall, M. (2005). Graphical views of suppression and multicollinearity in multiple linear regression. *American Statistician, 59,* 127–136.

Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 267–283.

Gerhand, S., & Barry, C. (1999). Age-of-acquisition and frequency effects in speeded word naming. *Cognition, 73*(2), B27–B36.

Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from building 20.* Cambridge, MA: MIT Press.

Hartshorne, J. K., & Ullman, M. T. (2006). Why girls say "holded" more than boys. *Developmental Science, 9,* 21–32.

Holmes, V. M., & Langford, J. (1976). Comprehension and recall of abstract and concrete sentences. *Journal of Verbal Learning and Verbal Behavior, 15,* 559–566.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Erlbaum.

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology, 104,* 130–136.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences, USA, 96*(13), 7592–7597.

Jones, G. V. (1985). Deep dyslexia, imageability, and ease of predication. *Brain and Language, 24,* 1–19.

Kennet, J., McGuire, L., Willis, S. L., & Schaie, K. W. (2000). Memorability functions in verbal memory: A longitudinal approach. *Experimental Aging Research, 26,* 121–137.

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language, 47,* 145–171.

Kieras, D. (1978). Beyond pictures and words: Alternative information-processing models for imagery effects in verbal memory. *Psychological Bulletin, 85,* 532–554.

Kimura, D. (1996). Sex, sexual orientation and sex hormones influence human cognitive function. *Current Opinion in Neurobiology, 6,* 259–263.

Kimura, D. (1999). *Sex and cognition.* Cambridge, MA: MIT Press.

Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 804–823.

Kramer, J. H., Delis, D. C., & Daniel, M. (1988). Sex differences in verbal learning. *Journal of Clinical Psychology, 44,* 907–915.

Kroll, J. F., & Merves, J. S. (1986). Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 92–107.

Kuczaj, S. A., II. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior, 16,* 589–600.

Lavric, A., Pizzagalli, D., Forstmeier, S., & Rippon, G. (2001). Mapping dissociations in verb morphology. *Trends in Cognitive Sciences, 5,* 301–308.

Lehtonen, M., Niska, H., Wande, E., Niemi, J., & Laine, M. (2006). Recognition of inflected words in a morphologically limited language: Frequency effects in monolinguals and bilinguals. *Journal of Psycholinguistic Research, 35,* 121–146.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation.* Chicago: University of Chicago Press.

Lewin, C., Wolgers, G., & Herlitz, A. (2001). Sex differences favoring women in verbal but not in visuospatial episodic memory. *Neuropsychology, 15,* 165–173.

Maitland, S. B., Herlitz, A., Nyberg, L., Backman, L., & Nilsson, L. G. (2004). Selective sex differences in declarative memory. *Memory & Cognition, 32,* 1160–1169.

Marchman, V. A. (1997). Children's productivity in the English past tense: The role of frequency, phonology, and neighborhood structure. *Cognitive Science, 21,* 283–304.

Marslen-Wilson, W., & Tyler, L. K. (1998). Rules, representations, and the English past tense. *Trends in Cognitive Sciences, 2,* 428–435.

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences, 6,* 465–472.

Mellet, E., Tzourio, N., Denis, M., & Mazoyer, B. (1998). Cortical anatomy of mental imagery of concrete nouns based on their dictionary definition. *NeuroReport, 9,* 803–808.

Meunier, F., & Segui, J. (1999). Morphological priming effect: The role of surface frequency. *Brain and Language, 68,* 54–60.

Miller, G. A. (1990). WordNet: An online lexical database. *International Journal of Lexicography, 3,* 235–312.

Moeser, S. D. (1974). Memory for meaning and wording in concrete and abstract sentences. *Journal of Verbal Learning and Verbal Behavior, 13,* 682–697.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 116–133.

Paivio, A. (1963). Learning of adjective–noun paired associates as a function of adjective–noun word order and noun abstractness. *Canadian Journal of Psychology, 17,* 370–379.

Paivio, A. (1965). Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior, 4,* 32–38.

Paivio, A. (1967). Paired-associate learning and free recall of nouns as a function of concreteness, specificity, imagery, and meaningfulness. *Psychological Reports, 20,* 239–245.

Paivio, A. (1971). *Imagery and verbal processes.* London: Holt, Rinehart, & Winston.

Paivio, A. (1995). Imagery and memory. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 977–986). Cambridge, MA: MIT Press.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs 76*(3, Pt. 2).

Pinker, S. (1999). *Words and rules: The ingredients of language.* New York: Basic Books.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28,* 73–193.

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

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., & Snyder, W. (1990, November). *Some evidence that irregular forms are retrieved from memory but regular forms are rule-generated.* Paper presented at the 31st Annual Meeting of the Psychonomics Society, New Orleans, LA.

Rubenstein, H., Garfield, L., & Milliken, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9,* 487–492.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & P. D. P. Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructures of cognition: Vol. 2. Psychological and biological models* (pp. 272–326). Cambridge, MA: Bradford/MIT Press.

Sabsevitz, D. S., Medler, D. A., Seidenberg, M., & Binder, J. R. (2005). Modulation of the semantic system by word imageability. *NeuroImage, 27,* 188–200.

Schwanenflugel, P. J., Akin, C., & Luh, W.-M. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition, 20,* 96–104.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 82–102.

Seidenberg, M. (1992). Connectionism without tears. In S. Davis (Ed.), *Connectionism: Theory and practice* (pp. 84–137). New York: Oxford University Press.

Seidenberg, M., & Bruck, M. (1990, November). *Consistency effects in the generation of past tense morphology.* Paper presented at the 31st Annual Meeting of the Psychonomic Society, New Orleans, LA.

Sereno, J. A., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory & Cognition, 25,* 425–437.

Sherwin, B. B. (2003). Estrogen and cognitive functioning in women. *Endocrine Reviews, 24,* 133–151.

Stemberger, J. P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition, 14,* 17–26.

Stemberger, J. P., & MacWhinney, B. (1988). Are inflected forms stored in the lexicon? In M. Hammond & M. Noonan (Eds.), *Theoretical morphology: Approaches in modern linguistics* (pp. 101–116). New York: Academic Press.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1140–1154.

Strain, E., Patterson, K., & Seidenberg, M. S. (2002). Theories of word naming interact with spelling–sound consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 207–214.

Tabak, W. M., Schreuder, R., & Baayen, R. H. (2005). Lexical statistics and lexical processing: Semantic density, information complexity, sex, and irregularity in Dutch. In S. Kepser & M. Reis (Eds.), *Linguistic evidence: Empirical, theoretical and computational perspectives* (pp. 529–555). Berlin: Mouton de Gruyer.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition, 7,* 263–272.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms.* New York: Wiley.

Trahan, D. E., & Quintana, J. W. (1990). Analysis of gender effects upon verbal and visual memory performance in adults. *Archives of Clinical Neuropsychology, 5,* 325–334.

Tyler, L. K., de Mornay-Davies, P., Anokhina, R., Longworth, C., Randall, B., & Marslen-Wilson, W. D. (2002). Dissociations in processing past tense morphology: Neuropathology and behavioral studies. *Journal of Cognitive Neuroscience, 14,* 79–94.

Ullman, M. T. (1993). *The computation of inflectional morphology.* Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Ullman, M. T. (1999). Acceptability ratings of regular and irregular past tense forms: Evidence for a dual-system model of language from word frequency and phonological neighbourhood effects. *Language and Cognitive Processes, 14,* 47–67.

Ullman, M. T. (2001a). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research, 30,* 37–69.

Ullman, M. T. (2001b). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience, 2,* 717–726.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92,* 231–270.

Ullman, M. T. (2005a). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods,* *theory and practice* (pp. 141–178). Washington, DC: Georgetown University Press.

Ullman, M. T. (2005b). More is sometimes more: Redundant mechanisms in the mind and brain. *APS Observer, 18,* pp. 4, 46.

Ullman, M. T. (2007). The biocognition of the mental lexicon. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 267–286). Oxford, England: Oxford University Press.

Ullman, M. T., Miranda, R. A., & Travers, M. L. (2008). Sex differences in the neurocognition of language. In J. B. Becker, K. J. Berkley, N. Geary, E. Hampson, J. Herman, & E. Young (Eds.), *Sex on the brain: From genes to behavior* (pp. 291–309). New York: Oxford University Press.

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

Ullman, M. T., Walenski, M., Prado, E. L., Ozawa, K., & Steinhauer, K. (under revision). *The compositionality and storage of inflected forms: Evidence from working memory effects.* Manuscript submitted for publication.

van der Lely, H. K. J., & Ullman, M. T. (2001). Past tense morphology in specifically language impaired and normally developing children. *Language and Cognitive Processes, 16,* 177–217.

Walenski, M., Mostofsky, S. H., Larson, J. C. G., & Ullman, M. T. (2008). Brief report: Enhanced picture naming in autism. *Journal of Autism Developmental Disorders, 38,* 1395–1399.

Walenski, M., Mostofsky, S. H., & Ullman, M. T. (2007). Speeded processing of grammar and tool knowledge in Tourette's syndrome. *Neuropsychologia, 45,* 2447–2460.

Walenski, M., & Ullman, M. T. (2005). The science of language. *Linguistic Review, 22,* 327–346.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48,* 817–838.

Yarmey, A. D., & O'Neil, B. J. (1969). S–R and R–S paired-associate learning as a function of concreteness, imagery, specificity, and association value. *Journal of Psychology, 71,* 95–109.

Received July 18, 2006
Revision received September 25, 2008
Accepted January 9, 2009 ■