In Press, *Journal of Memory and Language.*

Age of Acquisition Effects in Word Reading and Other Tasks

Jason D. Zevin
University of Southern California

Mark S. Seidenberg
University of Wisconsin - Madison

Recent studies have suggested that age of acquisition (AoA) has an impact on skilled reading independent of factors such as frequency. This result raises questions about previous studies in which AoA was not controlled, and about current theories in which it is not addressed. Analyses of the materials used in previous studies suggest that the observed AoA effects may have been due to other factors. We also found little evidence for an AoA effect in computational models of reading which used words that exhibit normal spelling-sound regularities. An AoA effect was observed, however, in a model in which early and late learned words did not overlap in terms of orthography or phonology. The results suggest that, with other correlated properties of stimuli controlled, AoA effects occur when what is learned about early patterns does not carry over to later ones. This condition is not characteristic of learning spelling-sound mappings but may be relevant to tasks such as learning the names for objects.

KEYWORDS: age of acquisition, reading, connectionist modeling, linguistic development

Many studies of word reading have examined how stimulus properties such as frequency, length, spelling-sound consistency, and imageability affect performance (see Balota, 1994; Seidenberg, 1995, for reviews). Over the past several years another factor, age of acquisition (AoA), has drawn considerable attention (Morrison & Ellis, 1995; Gerhand & Barry, 1998, 1999b, 1999a). The basic idea is that the age at which a word is learned in acquiring spoken language affects the performance of skilled readers. People learn words such as TOP and SYRUP before words such as TAX and SYRAH. As operationalized in recent studies, the AoA hypothesis is that there will be an effect of this early learning on adult performance when other factors such as frequency of usage in adult language are controlled.

Research supported by NIMH grant PO1-MH47566, NICHD grant RO1-MH 29891, and an NIMH research scientist development award to MSS. We thank Jay McClelland and Matt Lambon Ralph for helpful discussions. The models described in this article were implemented using software developed by Michael Harm, whom we also thank. Address for correspondence: Mark S. Seidenberg, Department of Psychology, University of Wisconsin, Madison WI 53706. Electronic mail address: marks@gizmo.usc.edu.

The existence of an AoA effect on word reading would be consistent with evidence concerning other types of age-dependent learning (Doupe & Kuhl, 1999; Quartz & Sejnowski, 1997). In many cognitive domains, early learning results in a reduction in plasticity that limits the ability to acquire new information. Phonological acquisition provides a classic example (Werker & Tees, 1984): learning the phonological structure of one's language limits the ability to learn new phonetic contrasts (e.g., in a second language). Similarly, there is evidence that the ability to learn the morphology and syntax of a language drops monotonically after approximately seven years of age (although it is controversial; see Flege, Yeni-Komshian, & Liu, 1999). Lexical acquisition is not thought to be highly age-dependent (Markson & Bloom, 1997; McCandliss, Posner, & Givon, 1997); still it is possible that early-learned words have an advantage over later-learned words, and that this would carry over to how they are read.

The ages at which people learned particular words are unknown, of course, but can be estimated from other measures. For example, Gilhooly and Logie (1980) collected subjective ratings of AoA, familiarity, imageability and concreteness for nearly two thousand words. These norms have been widely used in studies of effects of AoA on several tasks including tachistoscopic identification (Lyons, Teer, & Rubenstein, 1978), word naming (Brown & Watson, 1987; Coltheart, Laxon, & Keating, 1988) and object naming (Carroll & White, 1973; Ellis & Morrison, 1998) and with neurologically impaired patients (Hirsh & Ellis, 1994; Hodgson & Ellis, 1998; Lambon Ralph, Graham, Ellis, & Hodges, 1998). The Gilhooly and Logie (1980) data were obtained from 36 adult subjects; the AoA ratings also correlate significantly with independent measures of AoA (Gilhooly & Gilhooly, 1980; Lyons et al., 1978; Morrison, Ellis, & Chappell, 1997) suggesting that they provide reliable information.

Given estimates of the frequencies with which words occur in adult usage and when words were acquired, it seems natural to consider whether the two factors have independent effects on skilled performance. Morrison and Ellis (1995) orthogonally manipulated AoA and frequency in naming and lexical decision tasks, and found a strong AoA effect with frequency controlled, but no frequency effect with AoA controlled. They also observed that AoA and frequency had been confounded in previous studies, raising the possibility that effects attributed to frequency might have been due to AoA. Subsequent studies (Gerhand & Barry, 1998, 1999a, 1999b) replicated Morrison and Ellis' AoA effect with frequency controlled, but contrary to the earlier results, significant effects of frequency were observed with AoA controlled. Nonetheless, the finding that AoA affects performance independent of frequency seems to present a challenge for models of word reading (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989) that do not explicitly take this factor into account.

The research described below was motivated by empirical and theoretical considerations that led us to examine more closely whether age of acquisition has an effect on skilled reading. On the empirical side, the concern was that it might be difficult to isolate effects of age of acquisition because it is correlated with many stimulus properties, including frequency. Below we present analyses of the materials used in previous studies and other data which suggest that the evidence for an effect of AoA on skilled reading is weak at best. On the theoretical side, we were interested in developing a better account of why age of acquisition could have an effect on skilled reading or other tasks. Many previous studies have employed a bottom-up strategy in which AoA is treated as a factor, like frequency or length, that might account for independent variance in adult performance. However, AoA needs to be understood in terms of a theory that addresses why some words are learned earlier than others, and how early experience affects later performance. Such a theory would clarify the relationship between the AoA measure and other factors that affect word learning and skilled performance, and provide a stronger basis for generating predictions about the role of age of acquisition in reading and other tasks.

After examining existing studies of AoA effects in reading, we describe investigations of these effects using a computational model of the mapping from orthography to phonology (Harm & Seidenberg, 1999). Modeling was useful for several reasons. First, it allows direct manipulations of the frequency and timing of exposures to words using stimuli that are exactly controlled with respect to properties (such as frequency and length) that are normally highly confounded. Second, such models embody an explicit theory of reading acquisition and skilled processing in which the roles of frequency and timing of exposure can be examined. Finally, previous analyses of the behavior of such models suggest a possible computational basis for age of acquisition effects. In some models, the "entrenchment" of early-learned items has an effect on later performance (Ellis & Lambon Ralph, 2000; Munro, 1986). Thus, connectionist models are consistent with the existence of age of acquisition effects; our research addresses the conditions under which such effects occur and how they relate to the conditions that govern reading. We focused on the mapping between orthography and phonology because it plays an important role in the naming and lexical decision tasks that have been used to study AoA effects in reading.

To foreshadow the results, the simulations yielded two complementary findings. Simulations using a large corpus of English words yielded no effects of AoA on skilled performance. There was an initial advantage for words that were presented more often early in training, but there was no residual effect on skilled performance. This occurred because the regularities in the mapping between orthography and phonology that exist across words in English reduce the effects of early exposure to individual items. These results, taken with the analyses of previous behavioral studies, suggest that age of acquisition effects in word reading are likely to be minimal, with other properties that are correlated with AoA controlled. However, a significant age of acquisition effect was observed in a simulation in which early and late learned words were chosen so that they overlapped little in terms of orthographic or phonological structure. This artificial condition, which is not characteristic of reading acquisition, yielded an advantage for early-learned words in skilled performance with other factors controlled.

The simulations suggest that the occurrence of age of acquisition effects depends on the nature of the learning task, specifically whether what is learned about one pattern carries over to others with which it shares structure. Thus, we observed the effect in a simulation using materials that explicitly eliminated the overlap between early and late-learned patterns, but not when the stimulus patterns exhibited the regularities in the correspondences between spelling and sound that are characteristic of the English writing system. This analysis also extends to the simulations reported by Ellis and Lambon Ralph (2000), Smith, Cottrell, and Anderson (2001), and Monaghan and Ellis (in press), who observed robust age of acquisition effects using materials and tasks that differ from reading in important respects, discussed below. Thus both the modeling and the analysis of existing behavioral studies suggest that age of acquisition has little impact on skilled reading. At the same time, the modeling also suggests that such effects may occur for other tasks such as learning the names associated with objects or faces, for which the learning of one pattern carries little information about others. The full range of effects can be explained in terms of basic properties of learning in connectionist networks employing distributed representations. Such networks provide deeper insight about how early experience affects later performance.

Previous Studies

Two strategies have been used in previous studies of AoA effects in word reading. One is to conduct experiments in which AoA and frequency are manipulated factorially. The other is to use multiple regression to show that AoA accounts for unique variance in predicting reponse latencies or proportions of errors. We consider these in turn.

Morrison and Ellis (1995) conducted the first experiments factorically manipulating AoA and frequency in word reading tasks. Their stimuli were equated across conditions in terms of mean Kučera and Francis (1967) frequency, and other variables (e.g., imageability, length in letters, the N measure (Coltheart, Davelaar, Jonasson, & Besner, 1977)) but varied significantly in terms of rated AoA. This study and subsequent ones using similar methods (Gerhand & Barry, 1999b, 1999a, 1998; Monaghan & Ellis, in press; Turner, Valentine, & Ellis, 1998) yielded effects of AoA with such stimuli.

Table 1: Properties of the Stimuli Used in Previous Studies of Effects of Age of Acquisition and Frequency

Study	Condition	KF	log(KF)	Celex	log(Celex)	WFG	log(WFG)	FAM
Morrison &	Early	23	2.63	512	5.78	477	5.62	5.62
Ellis (1995)	Late	24	2.63	301	4.82	107	3.32	4.10
	Difference	-1	0	211	.96**	370**	2.30***	1.52***
Gerhand &	Early	105	3.01	1986	5.91	2164	5.41	5.35
Barry	Late	75	3.15	881	5.50	306	3.61	4.62
(1998,1999a,1999b)	Difference	30	-.14	1105	.41	1858†	1.80*	.73**
Turner et al.	Early	52	3.24	555	5.51	2184	6.90	5.69
(1998)	Late	50	2.86	309	4.63	1274	6.13	4.97
	Difference	2	.38	246	0.88**	910	0.77*	0.72***
Monaghan &	Early	35	2.63	654	5.56	411	5.20	NA
Ellis (in press)	Late	25	2.30	420	4.88	141	3.36	NA
Inconsistent Words	Difference	10	.33	234	.68	270*	1.84**	NA
Monaghan &	Early	33	2.14	672	4.97	469	4.31	4.97
Ellis (in press)	Late	29	2.07	496	4.93	199	3.76	4.55
Consistent Words	Difference	4	.07	176	.03	270	.65	.42

Note : In all cases, stimuli were matched using Kučera and Francis (1967). Turner et al. (1998) also matched their items on spoken frequencies from Baayen, Piepenbrock, and van Rijn (1993). WFG = Zeno (1995); FK =Kučera and Francis (1967); Celex = written English frequencies from Baayen et al. (1993); FAM = Familiarity from Gilhooly and Logie (1980). † $= p < .1$ * $= p < .05$; ** $= p < .01$; *** $= p < .001$. NA = Familiarity ratings were not available for most the Inconsistent items in Monaghan and Ellis (in press).

These studies raise concerns about whether stimulus frequencies were equated across conditions as the designs of these experiments required. Properties of words such as length in letters are objective and therefore easy to manipulate or control across conditions. In contrast, the frequency counts derived from corpora such as Kucera and Francis (1967) are statistics: estimates of a variable (how often a word is used) whose actual values are unknown. Like other statistics, frequency counts are associated with measurement error arising from factors such as the size of the corpus, the sample of texts used in generating the corpus, and individual differences in language experience. These sources of error can complicate the interpretation of frequency effects in behavioral studies

(Gernsbacher, 1984).

One problem is that the widely-used Brown corpus (from which the Kučera & Francis, 1967, norms are derived) is relatively small, which introduces considerable error in the estimates for individual words, particulary in the lower frequency range. Table 1 provides frequency data for the stimuli used in previous age of acquisition studies dervied from Kučera and Francis (1967) and two other sources, the Educator's Word Frequency Guide (WFG; Zeno, 1995) and Celex (Baayen et al., 1993) databases. Whereas the Brown corpus is about 1 million words, the WFG and Celex corpora are both over 16 million words. The data also include a measure of rated familiarity (Gilhooly & Logie, 1980), which Gernsbacher (1984) showed provides a more sensitive measure of frequency differences among lower frequency words. Morrison and Ellis (1995) equated their early and late AoA stimuli in terms of Kučera and Francis (1967) frequency, but as the table indicates, the items differ significantly on the other measures in the expected direction: early acquired words are also more frequent and familiar. The early and late stimuli in the Gerhand and Barry studies exhibit a similar pattern; there are numerical differences between the early and late stimuli on all measures, and they are significant using log WFG frequency and familiarity. The materials in the Turner et al. (1998) study also differ such that early words were higher in frequency (log Celex, log WFG) and rated familiarity than late words. In a recent study, Monaghan and Ellis (in press) examined age of acquisition effects for words with consistent or inconsistent spelling-sound correspondences. They equated the stimuli with respect to frequency estimates derived from both the Brown and Celex corpora. The stimuli in the inconsistent condition exhibit small differences in the direction of early words being higher in frequency on all three measures; using the WFG norms the difference is statistically reliable. For the consistent items, the differences between the conditions are smaller and nonsignificant on all three measures. The consistent condition is the only one in the table in which an age of acquisition effect was not obtained.

These cases are similar to the ones studied by Gernsbacher (1984), who showed that several apparently conflicting findings in the contemporary word recognition literature could be traced to the relative insensitivity of the Kucera and Francis frequency norms; stimuli that were apparently equated on this measure differed in terms of rated familiarity. In the studies in Table 1, stimuli that were equated on the Kučera and Francis (1967) norms differed in rated familiarity and/or another measure of frequency based on a larger corpus. The inconsistent word condition in the Monaghan and Ellis study is the least clear case, insofar as the stimuli did not differ reliably on two frequency measures but did on a third. It should be noted that the WFG norms appear to provide a sensitive measure of frequency, however. Table 2 presents the correlations among several measures of frequency and the naming and lexical decision latencies in three large-scale studies. The Seidenberg and Waters (1989) dataset consists of mean naming latencies for 3000 words from 30 undergraduate subjects; the Spieler and Balota (1997) data are naming latencies for 2,906 words from 31 subjects, and the Balota, Pilotti, and Cortese (2001) data are lexical decision latencies for 2,905 words from 60 subjects (30 young adults and 30 older adults). The correlations between estimated frequencies and response latencies are highest for the WFG norms, which also account for unique variance when entered into a simultaneous multiple regression with the other norms. Below we return to methodological issues about the use of different frequency norms; here the main point is that the early and late acquired stimuli in previous studies were not closely matched in frequency and thus did not provide strong tests of the role of age of acquisition independent of this factor. [1]

[1]Another bit of evidence that the age of acquisition effect reported by Monaghan and Ellis (in press) was due to differences in frequency is reported by Strain, Patterson and Seidenberg (submitted), who found that using frequency

Table 2: Various Frequency Measures as Predictors of Naming Latency in Large-Scale Studies

Study	Measure	r	Unique Variance (%)
Spieler and Balota, 1997	WFG	-.35	2.39***
	FAM	-.32	.82*
	CELEX	-.29	.12
	KF	-.27	.03
Seidenberg and Waters, 1989	WFG	-.23	.72*
	FAM	-.21	.22
	CELEX	-.21	.11
	KF	-.18	.27
Balota, Pilotti and Cortese, submitted	WFG	-.63	3.97***
	FAM	-.62	3.86***
	CELEX	-.58	.22
	KF	-.51	.80**

Note: ** $= p < .01$; *** $= p < .001$. WFG = Word frequency from Zeno (1995), FAM = familiarity from Gilhooly and Logie (1980), CELEX = frequency from Baayen et al. (1993), KF = frequency from Kučera and Francis (1967).

Some of the studies in Table 1 also included conditions in which age of acquisition was controlled and frequency varied, which yielded a mixed pattern of results. Morrison and Ellis (1995) found a frequency effect in lexical decision, but not in naming; age of acquisition effects, in contrast, were found in both tasks. The fact that there was an AoA effect but not a frequency effect in the naming task suggested that the AoA effect could not be wholly due to a frequency confound. However, this pattern of results did not replicate in a study by Gerhand and Barry (1998) using the same stimuli; they observed both frequency and age of acquisition effects in naming. The Morrison and Ellis (1995) data also exhibited an atypical pattern in which lexical decision latencies were faster than naming latencies for the same words (cf. Balota & Chumbley, 1984; Forster & Chambers, 1973). In summary, the factorial studies leave open a window of uncertainty as to whether the observed effects were due to differences in age of acquisition or frequency.

The second methdology employed in this area involves using multiple regression to isolate unique variance in response latencies associated with AoA (Brown & Watson, 1987; Butler & Hains, 1979; Lyons et al., 1978; Morrison & Ellis, 2000). These studies reported effects of AoA independent of other stimulus properties including imageability, familiarity and frequency. We conducted a similar analysis using the data from the three large-scale studies of word naming and lexical decision mentioned above (Seidenberg & Waters, 1989; Spieler & Balota, 1997; Balota et al., 2001) and found similar results. For 528 of the words in these studies, there are data concerning both frequency (Zeno, 1995) and AoA (Gilhooly & Logie, 1980). For all three data sets, AoA and frequency were significantly correlated with response latencies (Table 3); for the Spieler and Balota (1997) and Balota et al. (2001) data both factors account for unique variance.

It is important to avoid making a “correlation is causation” error in interpreting these data,

counts derived from either the Celex or WFG databases as a covariate in the analyses of variance eliminated the age of acquisition effect in the Monaghan and Ellis (in press) data.

Table 3: Frequency and Age of Acquisition as Predictors of Naming Latencies

Study	Measure	r	Unique Variance (%)
Spieler and	WFG	-.28	2.59***
Balota, 1997	AoA	.28	2.35***
Seidenberg and	WFG	-.19	1.52**
Waters, 1989	AoA	.17	.64†
Balota, Pilotti	WFG	-.49	9.20***
and Cortese, submitted	AoA	.44	5.15***

Note: †= $p < .1$; ** = $p < .01$; *** = $p < .001$. WFG = Word frequency from Zeno (1995)

Table 4: Correlations Among 6 Standard Lexical Measures and AoA

Variable	AoA	WFG	IM	FAM	CON	LEN
WFG	-0.5141***					
IM	-0.5861***	0.1073*				
FAM	-0.6740***	0.7203***	0.2026***			
CON	-0.3840***	0.0056	0.8082***	-0.0099		
LEN	0.1984***	-0.0666	-0.1483***	-0.0605	-0.1717***	
N	-0.1976***	0.1417**	0.1195**	0.1245**	0.1215**	-0.7142***

Note: * $= p < .05, ** = p < .01, *** = p < .001$. WFG = log Zeno (1995) frequency; IM = imageability; FAM = familiarity (Gilhooly & Logie, 1980); CON = concreteness; LEN = number of letters; N = Coltheart's N.

however, because both AoA and frequency are correlated with other stimulus properties. To illustrate, Table 4 provides the correlations among AoA, frequency, Coltheart's N, length in letters, and rated familiarity, imageability, and concreteness (also from the Gilhooly & Logie, 1980, norms) for the 528 words. These intercorrelations make it difficult to isolate effects due to age of acquisition per se. Some additional information is provided by assessing the amount of unique variance associated with frequency and age of acquisition after the other measures in Table 4 have been partialled out (Table 5). These results indicate that whereas frequency accounts for a small but significant amount of variance, the age of acquisition measure does not[2]. These data suggest that, rather than there being an effect of age of acquisition on skilled performance independent of other stimulus factors, the ages at which words are learned are determined by factors such as frequency, length, and imageability. Thus, after these factors are taken into account, there is no residual effect associated with the age of acquisition measure.

The results in Table 5 differ from those reported by Brown and Watson (1987) and Morrison et al. (1997), who conducted similar analyses using smaller sets of words and found significant

[2]The amount of unique variance attributed to either variable is surprisingly small. One factor that may be relevant is that effects of lexical frequency are reduced or eliminated by exposure to neighboring words. Words that have many neighbors (e.g., consistent ones) do not show strong frequency effects in naming. Another is that naming is less sensitive to frequency effects than other tasks because it only measures time to initiate the response; frequency effects can also show up in things like duration of the whole utterance (Balota & Abrams, 1995) and in the duration of onsets that contain continuants (Kawamoto, Kello, Jones, & Bame, 1998).

Table 5: Unique Variance Accounted for by Frequency and AoA Independent of Other Lexical Variables

Study	Measure	Unique Variance (%)
Spieler and	WFG	1.27**
Balota, 1997	AoA	.29
Seidenberg and	WFG	.69*
Waters, 1989	AoA	.01
Balota, Pilotti	WFG	2.94***
and Cortese, submitted	AoA	.34

Note: *** = p < .001; ** = p < .01; * = p < .05; WFG = cumulative frequency from Zeno (1995), AoA = age of acquisition from Gilhooly and Logie (1980)

effects of age of acquisition independent of frequency. The differing results appear to be related to differences between the WFG norms and the Brown and CELEX norms used in earlier studies. The WFG norms are based on a larger sample of texts than the Brown norms and the sample is more diverse than either the Brown or Celex samples. Like the American Heritage norms (Carroll et al., 1971), the WFG sample includes texts from a broad range of reading levels, including books for school-aged children. Each text in the sample was assigned a grade-level based on a readability formula. Frequency data are provided for each word at each grade level, ranging from first grade to college. For the analyses presented above, we used the sum of these frequencies. The fact that the WFG frequencies correlate more highly with response latencies than the other norms (Table 2) and yield no residual effect of age of acquisition (Table 5) may be related to the inclusion of this broader range of texts.

Table 6: Unique Variance Accounted for by AoA with Different Subsections of the WFG Norms Used as Predictors

		WFG Subsection							
Study	Predictor	2-13+	3-13+	4-13+	5-13+	6-13+	7-13+	8-13+	9-13+
SB	AoA	.36	.41	.44†	.47†	.50†	.54†	.56†	.57†
	Frequency	1.26**	1.17**	1.01**	.85*	.84*	.86*	.78*	.67*
SW	AoA	.04	.04	.06	.07	.08	.10	.10	.12
	Frequency	.98*	.97*	.89*	.83*	.87*	.95*	.91*	.91*
BCP	AoA	.39†	.46†	.52*	.58*	.63*	.68*	.72*	.68*
	Frequency	2.43***	2.22***	2.04***	1.92***	1.97***	2.10***	2.11***	2.18**

Note: †= p < .10; * = p <.05; ** = p < .01; *** = p < .001; WFG = Zeno (1995) frequency counts; 2-13 = Grade levels 2 (2nd grade) to 13+ (University) in the WFG norms. SB = Spieler and Balota (1997); SW = Seidenberg and Waters (1989); BCP = Balota et al. (2001)

To examine this issue further, we conducted regression analyses using different subsets of the

Table 7: Correlation Between AoA and WFG Frequency at Different Grade Levels

Grade Level													
1	2	3	4	5	6	7	8	9	10	11	12	13	TOTAL
-.68	-.67	-.63	-.60	-.53	-.50	-.47	-.45	-.43	-.38	-.35	-.31	-.17	-.51

Note: All correlations significant, $p < .001$.

WFG corpus. Specifically, we examined how much variance the WFG and AoA measures accounted for when the data from lower grades were excluded (Table 6). The results for all three of the large-scale behavioral studies exhibit a consistent pattern: as more of the data from lower grade-levels is excluded, the amount of residual variance due to frequency decreases while the amount associated with AoA increases. In two of the three studies, the AoA effect reaches significance with data from the younger grades excluded, although the amount of variance account for is very small.

One interpretation of these results is that there is a small effect of age of acquisition on skilled performance which the WFG norms (but not Brown or Celex) pick up because the corpus included texts for younger readers. Words that are learned earlier may tend to be used more often in texts that are appropriate for younger readers. Table 7 presents the correlations between rated age of acquisition and grade-level frequency for the 528 words used in previous analyses; there are strong negative correlations which decline gradually with age. Thus it could be argued that the WFG frequency data for the lower grades covertly encode age of acquisition. On this view, skilled performance is affected by two independent factors, age of acquisition and frequency of usage in adult language, both of which are captured by the cumulative WFG frequency measure.

There is a different explanation for these results, however: unlike the Brown or Celex corpora, the WFG norms provide estimates of the cumulative frequencies of words, that is, how often they have been encountered over a long period of time (e.g., since an individual began to read). Cumulative frequency may be a better predictor of adult performance because it affects how lexical information is represented in memory (as for example in the connectionist models discussed below). On this view, age of acquisition norms account for variance in skilled performance because they index how frequently words were used at younger ages, information that the Brown and Celex norms do not include. Thus there is an effect of cumulative frequency on skilled performance, rather than separate effects of age of acquisition and adult frequency of usage. The WFG norms provide a reliable estimate of cumulative frequency, leaving no residual effect of age of acquisition.[3]

In summary, the data in Table 1 and the correlational analyses suggest that the age of acquisition effects observed in previous studies may have been due to confounds with "adult" frequency (measured by Kucera & Francis and Celex) or cumulative frequency (assessed by WFG). One difficulty in developing a well-controlled AoA experiment arises from the strong correlations between AoA and other lexical variables presented in Table 4. These correlations make it difficult to design

[3]It is important to recognize that the grade-level frequency data in the WFG norms are not literally data concerning the grades (or ages) at which the texts were read. Rather, they reflect the assignment of texts to grade levels using a formula that weighs factors such as number of words per sentence and number of syllables per word. On this measure, *Charlotte's Web* and *The Old Man and the Sea* are both assigned to the 4th grade reading level, for example. Thus, the data from the lower grade-levels reflect texts that are likely to be read by children at a given age but also texts of approximately similar structural complexity that are read at older ages. On our view (supported by the modeling presented below), these norms are relevant because they provide estimates of the cumulative frequency, rather than the exact timing, of exposures to words.

factorial experiments in which AoA is varied for a sufficient number of items with these and other factors controlled. The regression analyses suggest that AoA may account for a small amount of variance in skilled performance because it is correlated with how often words are read at younger ages, data that are not indexed by "adult" norms such as Kucera and Francis (1967) but which contribute to cumulative frequency of exposure.

Theoretical Issues

The above discussion addressed some methodological issues that arise in attempting to isolate age of acquisition effects. The data indicate a need to consider what statistics such as estimated age of acquisition and frequency measure and how they relate to the mechanisms that underlie lexical acquisition and processing. The concept "age at which a word is acquired" seems clear enough and intuitively different from "frequency of usage in adult language." However, whereas frequency norms reflect a property of words (namely, how often they are used), age of acquisition norms reflect something different, a behavioral event (learning a word by a certain age). This event is very similar to a task such as naming aloud: one behavior concerns how long it took to learn a word, the other how long it takes to pronounce a word. This point is particularly clear with respect to "objective" measures of AoA (Morrison et al., 1997) obtained by determining the ages at which children can name pictured objects. Just as studies of word reading have examined the factors that make some words easier to name than others, age of acquisition can be considered with respect to the factors that cause some words to be learned earlier than others.

Among these factors is frequency. In many theories, the frequency with which a stimulus is practiced or experienced affects how early and well it is learned as well as skilled performance. If the age at which a word is learned is affected by how often it is experienced, empirical estimates of AoA may covertly encode frequency of occurrence during the acquisition period. Moveover, we have also seen that age of acquisition ratings are correlated with grade-level frequency data from the WFG norms, including data from higher grades well past the ages at which the words were acquired. Thus, age of acquisition norms appear to be related to frequency of occurrence over a multi-year time span beginning with initial acquisition.

Seen in this light, word frequency, as standardly operationalized using norms such as Kucera and Francis (1967), provides the remaining chronological data concerning how often words are experienced in adulthood. These observations suggest that both age of acquisition and "adult" frequency norms reflect how often words are encountered but at different points in a developmental continuum ranging from initial acquisition to adulthood. The WFG norms take matters one step further, providing estimates about how often words are encountered at multiple points along this continuum, as well as about cumulative frequency. Thus, age of acquisition and frequency seem more intrinsically related than recent discussions have suggested. In effect, studies like the ones in Table 1 attempted to dissociate the effects of frequency of exposure during two widely-spaced time spans.

Connectionist modeling

Connectionist models of reading that employ distributed representations and gradual learning from experience provide a theoretical framework for examining effects of the frequency and timing of learning experiences on performance (e.g., Harm & Seidenberg, 1999; Plaut et al., 1996; Seidenberg & McClelland, 1989). Such models illustrate three points relevant to the AoA hypothesis. First, frequency has pervasive effects on network performance, including how quickly a word is learned ("age of acquisition") and level of skilled performance. Second, these effects are intrinsically related. Models such as Seidenberg and McClelland's (1989) attempt to provide unified account of acquisition and skilled performance in which the same computational principles apply throughout the developmental continuum. The effects of frequency on learning a word and on skilled performance are both realized by changes to the weights governing network performance. Thus the behavior of the system reflects the cumulative effects of exposure to words over time. Finally, the magnitudes of the effects of frequency of exposure differ depending on the state of the network, which changes over time as knowledge is acquired. As the model picks up on the similiarities that hold across words, and as the weights assume values that allow output to be produced accurately (i.e., minimize error), the effects of pattern frequency decline.

Some properties of these networks favor the idea that there will be an advantage for words that are learned earlier in training (Ellis & Lambon Ralph, 2000). (We assume for the remainder of this discussion that stimuli are equated along other dimensions.) Consider a network such as Seidenberg and McClelland's in which weights are initially set to random values and output units take values of 1 or 0. The adjustments to the weights that occur using backpropagation with a logistic activation function are proportional to the activation of the unit according to the term $a(1 - a)$, where a is the activation value. The adjustments are therefore largest when the activations are in the middle of the logistic function (around .5), as occurs when the network is initialized with small, random weights. The adjustments become smaller as the weights assume values that cause unit activations to more closely approximate the target values of 1 or 0. Thus, there is a loss of plasticity associated with learning the early-trained patterns. In effect, early-trained patterns become entrenched in the weights (see Munro, 1986, for an early discussion of this phenomenon). Both Ellis and Lambon Ralph (2000) and Smith et al. (2001) emphasize these aspects of network behavior in explaining age of acquisition effects.

There is another factor to consider, however: the effects of similarities across training patterns. The mapping between spelling and sound in English exhibits considerable systematicity. Reading models such as Seidenberg and McClelland's employed representations that allowed the weights to encode these regularities. Thus what is learned about one word carries over to other words with which it shares structure. This property modulates the effects of exposure to a given word. Until the model begins to encode the systematic aspects of the mapping, performance on a pattern is highly dependent on how often it is trained. By later in training the weights reflect the structure of the entire training set, changing its behavior. Once a word is learned, additional repetitions have little impact, creating a discrepancy between frequency of training and network performance. Furthermore, new words can be learned with little training if they share structure with known words. In the limit a new word can be pronounced correctly with no training, as in nonword generalization. Thus, there is an initial advantage for words that are trained with high frequency, but as the model learns there is less and less of a disadvantage for later-trained items. In effect the entrenchment of early-learned words is reduced as the model picks up on patterns that hold across words (see also Marchman & Bates, 1994).

In summary, the entrenchment phenomenon in connectionist networks provides a basis for age of acquisition effects, but other properties of the task and materials to be learned will affect whether there is a long-lasting effect on performance, as the age of acquisition hypothesis suggests.

Using this theoretical framework, the issue of AoA effects in reading can be clarified by considering two factors, *cumulative frequency* and *frequency trajectory*. Cumulative frequency refers to how often a word is presented to the network from the beginning to the end of training. This is a simplified analogue of how often people have encountered a word to the point at which performance is assessed. Frequency trajectory refers to how experience with a word is distributed over time. Thus, a given cumulative frequency can be associated with different trajectories.

The AoA hypothesis, then, is the prediction that frequency trajectory has an effect on adult performance independent of cumulative frequency. Specifically, if the cumulative frequencies of words (as well as other stimulus properties) are equated, words for which most of the training occurs early should show an advantage over words with other trajectories. Words that are trained more often early in development will in general be learned earlier than words that are mainly trained later; thus frequency has an effect on age of acquisition. However, the age of acquisition hypothesis is that there will be a further effect of this early experience on skilled peformance.

A measure such as Kučera and Francis (1967) frequency provides a poor estimate of cumulative frequency. Given the nature of the texts used to generate the corpus, it tends to underestimate the frequencies of many low frequency words, including ones that are mainly experienced in childhood. The WFG norms probably provide better information about cumulative frequency, but this is difficult to independently assess. Age of acquisition norms, in contrast, provide imperfect information about frequency trajectory because some words that are learned early (e.g., BOTTLE, CUP) are also used frequently later in life whereas others (e.g., TEDDY, BOOTIE) are not.

Because the actual cumulative frequencies and frequency trajectories of different words are not known, and because frequency norms and rated AoA provide imperfect estimates, we took the approach of using simulation modeling to explore the phenomena. Simulation also allowed control over stimulus properties that are normally confounded. Thus we could create conditions in which it was certain that cumulative frequency and stimulus properties were closely matched, while manipulating frequency trajectory, providing a strong test of the age of acquisition hypothesis.

Simulation 1

In the first simulation, a model was trained on a large corpus of words using the standard technique of probabilistically presenting words during training as a function of their estimated frequencies of occurrence (Seidenberg & McClelland, 1989). The critical data concern a subset of items for which we manipulated frequency trajectory while keeping cumulative frequency constant. Some of these words were more frequent early in training compared to later (Early condition), whereas other words followed the complementary trajectory (Late condition). By the end of training, however, cumulative frequencies of words in the two conditions were the same. In addition, the same words appeared in both Early and Late conditions across different runs of the model.

This model differs from previous models of age of acquisition effects in an important way: the task was closely related to the problem of learning the spelling-sound correspondences of English, information that plays an important role in the naming and lexical decision tasks used in the behavioral studies discussed above. The input and output representations were based on English orthography and phonology and the training corpus, a large set of monosyllabic words, instantiated the quasiregular mappings between the two (Seidenberg & McClelland, 1989). Previous simulations have utilized more artificial tasks and stimuli that did not capture this rich structure (discussed further below). Simulation 1 therefore provides more direct evidence concerning the occurrence of age of acquisition effects in reading.

Methods

Architecture.

The basic architecture shown in Figure 1 was used in all simulations. For Simulations 1 and 2, models with 100 orthographic (input) units, 250 phonological (output) units and 100 hidden units were used. In addition, the phonological layer had 20 hidden units which mediated connections between this layer and itself (cleanup units; Hinton & Shallice, 1991) . The cleanup units diffentiate this model from a simple feedforward net such as the one studied by Seidenberg and McClelland (1989). The network is given an input pattern and activation spreads through the network over a series of time steps. Each unit propagates activation to the other units to which it is connected. The feedback connections between the phonological and cleanup units create a type of dynamical system called an attractor network which settles into a stable pattern over time (see Harm & Seidenberg, 1999, for additional details). A further feature of the model was that each time step was discretized into a series of moments, which allows a unit's activation to ramp up gradually. Thus the learning algorithm (continuous recurrent backpropagation) changes the weights in ways that improve accuracy but also how quickly the network produces the correct output (see Harm, 1998; Bishop, 1995, for discussion).

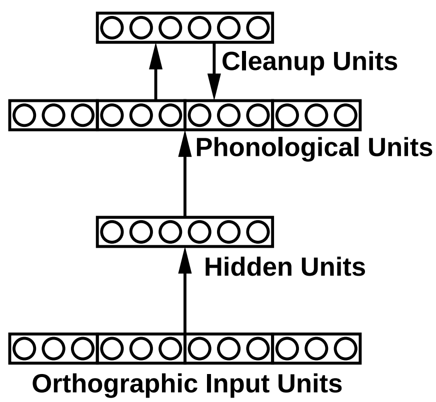

Figure 1. Model architecture used in all simulations

Corpus and Training. The training corpus consisted of 2,891 monosyllabic, monomorphemic words. 108 of these words were critical items whose frequencies were manipulated, as detailed below. The remaining 2,783 words (background items) were assigned frequencies taken from the Marcus, Santorini, and Marcinkiewicz (1993) norms, which are based on 43 million tokens from *The Wall Street Journal*.[4]

[4]The Wall Street Journal corpus has been extensively used in sentence processing research and at the time we began this research it was the largest available corpus of English. The lexical sample is somewhat skewed insofar as words such as STOCK, MARGIN, and INFLATION are overrepresented compared to other corpora. In our simulations, the norms were only used to insure that the background items in the training set were presented with a distribution of frequencies similar to that seen in natural language. When the goal is to examine the effects of frequency on individual words, other norms such as Zeno (1995) are preferable.

The critical items were divided into two lists of 54. Sets of 4 items were created by exchanging onsets and rimes. The lists were counterbalanced such that, for example, FOIST and MIST occurred on one list and FIST and MOIST on the other. Thus each list contained each onset and rime in the quadruple, but in different combinations. The model was run ten times with different initial random weights (between 0.1 and -0.1), analogous to replications with different subjects. Each list occurred five times in each trajectory. Thus the same items occurred in both Early and Late conditions across simulations. The data presented below are averages across the 10 runs of the model.

The Early and Late trajectories were designed to provide a strong test of the effects of early exposure on later performance; they were not intended to capture the observed trajectories for individual words, which are more variable. The frequencies of the words in the Early and Late conditions were manipulated as follows. Training consisted of ten epochs of 100,000 trials each. Early items were assigned a frequency of 1000 for the first three epochs of 100,000 training trials. For the next four epochs the frequency was adjusted to 500, 100, 50 and 10 in succession. Finally, for the last three epochs the frequency was set to one. The trajectory in the Late condition was the complement of the one in the Early condition. Late items started at a frequency of 1 for the first 3 epochs, frequency was adjusted to 10, 50, 100 and 500 over the next 4 epochs, and it finally reached 1000 for the last three epochs. These frequencies are within the range of the raw Marcus et al. (1993) frequencies used for the background items. As with the frequencies used for the non-critical words, these assigned frequencies were square-root transformed and items were sampled probabilistically. This method of compressing the frequency distribution allows the model to learn very low frequency items after a relatively small number of trials (Plaut et al., 1996). The actual frequencies with which the critical items were presented to the model at each epoch are given in Figure 2. The mean frequency for Early items in the first epoch was 41 and the mean frequency of Late items in this same epoch was 4. Frequencies were adjusted over time such that in the last epoch, the Late items had a mean frequency of 40 and the Early items had a mean frequency of 4. Importantly, by the end of training the Early and Late words had been trained equally often: the cumulative frequencies averaged across items were 198 for Early words and 196 for the Late words, $t(107) < 1$.

On each training trial, a word was probabilistically selected for training and its orthographic pattern was activated on the input units. Activation propagated forward for 11 time ticks. On the 12th time tick, error was computed and the weights of the model adjusted accordingly. The learning algorithm computes error on the basis of the difference between the desired and observed output at a given time tick, as well as the state of the model at earlier time ticks. In this way, each adjustment of the weights leads to incrementally more accurate as well as faster computation of the desired output.

Results and Discussion

The model's performance was assessed using both accuracy and sum squared error (SSE) measures. The model's output for a word was scored as correct if the output for each phoneme was closer to the correct phoneme than any other by euclidean distance. The SSE measure was the sum of the squared differences between the computed output and the target. The two measures are highly related; correct words produce lower error scores than incorrect words. However, among the correct words, differences in SSE reflect the relative difficulty of generating a response (see, e.g., Seidenberg & McClelland, 1989). Thus, the model's performance can continue to improve after it

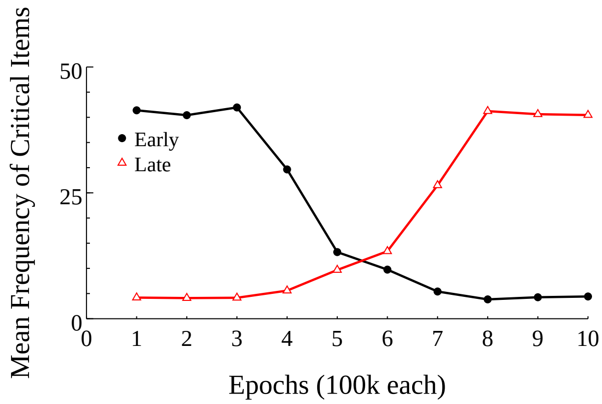

Figure 2. Frequency trajectories of critical items in Simulation 1

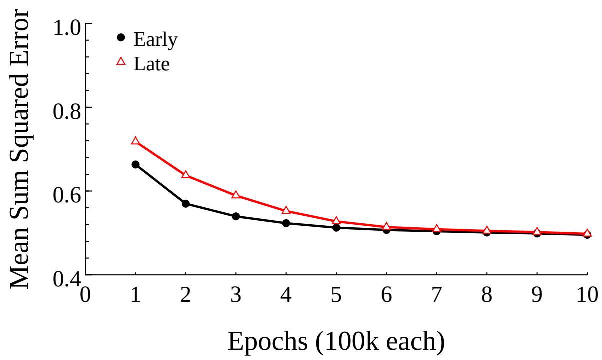

Figure 3. Performance over time for critical items in Simulation 1

has learned to produce the correct response, as in human performance.

At the end of training, the model produced correct output for 98% of the training set. Errors were almost all on low frequency strange words such as COUP, PLAID and RHEUM, which are thought to require input from the orthography → semantics → phonology pathway that was not implemented here (Plaut et al., 1996; Strain, Patterson, & Seidenberg, 1995; Harm & Seidenberg, 2001).

For the smaller set of critical words, the model learned to produce correct output for all items within the first epoch. Mean sum squared error for these items was calculated after each epoch. As shown in Figure 3, there was a small effect of frequency early in training which rapidly disappeared. T-tests on the difference between the means in the Early and Late conditions confirmed this: Error scores were significantly lower for Early words compared to Late after the first epoch, $t(107) = 4.24, p < .001$, and this effect remained significant after 5 epochs, $t(107) = 2.09, p < .05$. By epoch 6, when the frequency trajectories began to cross, the effect was nonsignificant, $t(107) = 1.12, p > .1$. At the end of training, when the cumulative frequency of the two groups was closely matched, there was also no reliable difference between conditions; in fact the means were identical, .50. At this point all critical items were still pronounced correctly.

The first simulation indicates that with stimulus properties equated, there is an effect of frequency trajectory early in training, but this effect rapidly recedes. By the end of training, when the cumulative frequencies are equated, there is no residual effect. Early in training, before much learning has occurred, performance is better on words that are trained more often. This is simply a frequency effect during the early phase. As training continues, performance in the two conditions converges to the same level.

Simulation 2

Simulation 2 was a replication of the first simulation that addressed two concerns. First, effects of the frequency trajectory manipulation might have been difficult to detect because the critical stimuli all contained spelling patterns with consistent spelling-sound correspondences. In addition, the stimuli were constructed in quadruples such as FIST-MOIST-MIST-FOIST, insuring that every word-body occurred at least twice with the same pronunciation. In the type of network studied here, learning of one item with a given spelling-sound pattern (e.g., FIST) carries over to other items containing the same pattern (e.g., MIST), reducing the effects of exposure to the item itself (a neighborhood effect). The net result was that all of the critical words were learned relatively rapidly; there was an effect of frequency of exposure early in training but it was observed on the sum squared error measure, not how rapidly the model learned (i.e., "age of acquisition"). We therefore created a new set of critical stimuli containing only "strange" words (Seidenberg, Waters, Barnes, & Tanenhaus, 1984) which have atypical spellings and spelling-sound correspondences. Because they have few close neighbors, these words show larger effects of frequency both in behavioral studies (e.g. Seidenberg et al., 1984) and connectionist models (e.g., Seidenberg & McClelland, 1989). We therefore expected to see effects of frequency trajectory on both SSE and how quickly these words were learned.

A second issue concerns the processes that gave rise to the Figure 3 data. One possibility is that these data reflect two complementary "age of acquisition" effects. Thus far we have followed the behavioral research in emphasizing the possible effect of early high frequency exposure on skilled performance. There might also be a complementary effect of high frequency exposure late in training, however. Thus the similar levels of performance in the Early and Late conditions at the end of training might derive from two sources: an AoA effect *and* a recency effect (Lewis, 1999, found evidence for both in a face naming task). We therefore added a control condition using a relatively flat frequency trajectory. For this condition, a subset of the critical items from Simulation 1 were assigned their normal frequencies and included among the background stimuli. After running the simulation, we isolated a large subset of these words that met two conditions: (a) their frequency trajectories were very flat, and (b) their cumulative frequencies were similar to what they were in Simulation 1. Thus the flat trajectory condition acts as a baseline against which the data from Simulation 1 can be compared. An effect of either the Early or Late trajectory in Simulation 1 would be indicated by better performance than in the flat trajectory condition at the end of training.

Finally, the flat trajectory condition was also used to assess whether cumulative frequency has an effect on network performance independent of trajectory, by comparing the results for two subsets of stimuli from the flat condition whose cumulative frequencies were considerably different.

Methods

The same model and corpus were used as in Simulation 1. The critical items from the earlier simulation were included among the background items and assigned their Marcus et al. (1993) frequencies, and a different set of 48 critical items was selected. The main criterion for the critical items was that their bodies not be assigned the same pronunciation in other words in the training list; thus, they included words such as BEIGE, PHLEGM and SCOURGE. The stimuli were divided into two lists with the assignment of lists to training condition counterbalanced across two simulations. The mean cumulative number of presentations for both Early and Late words was 183.

Stimuli in the Flat trajectory condition consisted of 95 of the critical stimuli in Simulation 1. These items were selected because when presented throughout training at their standard Marcus et al. (1993) frequency, they are well matched to the critical items for cumulative frequency. The mean cumulative frequency of these words was 200, comparable to the cumulative frequencies for these words in the Early and Late conditions in Simulation 1 (198 and 196, respectively).

Results

After 10 epochs, the model generated correct phonological codes for 98% of the training set. Performance on the critical items was assessed in terms of SSE, accuracy, and how quickly words were learned (i.e., "age of acquisition" in model time). Because the models were initialized with different random weights and because words were selected probabilistically during training, individual runs of the model differ slightly from one another in terms of performance, including when in training individual words were learned. Analogous individual differences are seen in children. For each item, age of acquisition was defined as the point at which 75% of the models generated correct responses. This criterion is similar to one used in the Morrison et al. (1997) study in which the age at which children acquired a word was defined as the age at which 75% of the subjects could name a pictured object accurately. By this measure, the average "age" at which Early items were acquired was approximately 2.09 epochs, whereas the average age for Late items was approximately 6.7 epochs. This difference is significant, $t(34) = 12.14$. Note that epochs are defined with respect to the total number of training training trials on all items, including the 2,843 background words, not the number of exposures to individual words. The mean number of trials to learn words in the Early and Late conditions were 296 and 250, respectively. These data indicate that the Early words were acquired more rapidly than the Late words, as expected. It took fewer exposures to learn the Late words because they benefitted from prior learning of other words. Even for strange words, then, there is generalization based on exposure to other words.

Accuracy over the course of training is depicted in Figure 4A. As in the previous simulation, the advantage for the early items dissipated as the cumulative frequency of the Late items converged on that for the Early items. Mean accuracy for both conditions was 85% at the end of training. This level of accuracy is somewhat lower than for the consistent words in Simulation 1; this finding is consistent with the view that performance on the most difficult strange words normally requires input from orthography → semantics → phonology. The error rate did not differ in the two frequency trajectory conditions, however, $t(47) < 1$. Thus, although the frequency trajectory manipulation affected the "age" at which items were acquired, it had no residual effect on accuracy when the cumulative frequency of Early and Late items converged. Figure 4B shows the change in sum squared over time for Early and Late items, which is very similar to the accuracy graph.

One further aspect of the data is worth noting: Toward the end of training the model be-

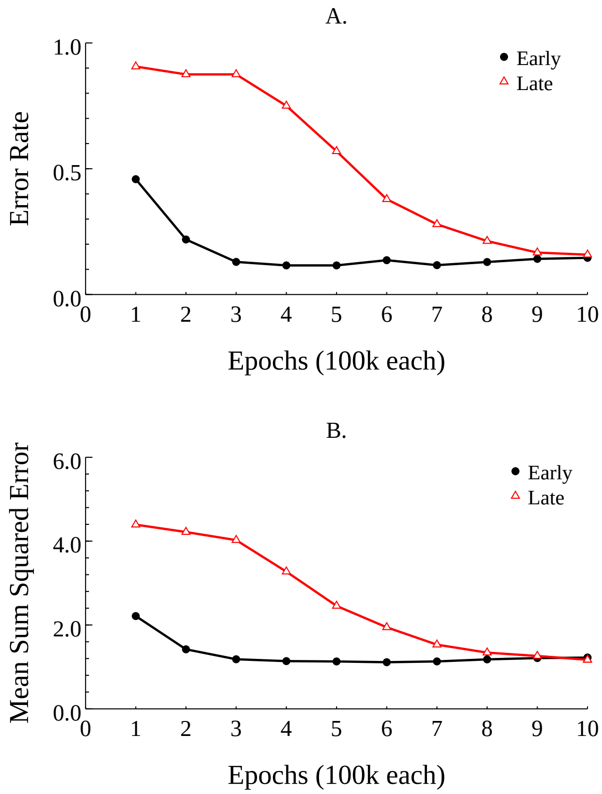

Figure 4. Performance over time for Simulation 2, A) error rate and B) sum squared error

gan to exhibit some unlearning of the Early words, as indicated by the slowly rising scores in this condition for both measures. Protecting early-acquired words from unlearning requires intermittent re-exposure to these items over time (Hetherington & Seidenberg, 1989). The Early trajectory entailed a steep decline in frequency toward the end of training. This property, taken with the probabilistic nature of sampling, resulted in too few exposures to maintain performance at the maximum level. We did not systematically examine performance after 10 epochs, because it was at this point that the two conditions converged on the same cumulative frequencies. We do know, however, that a small number of additional training trials on the critical items is sufficient to stop the slow erosion of performance seen in Figure 4. This behavior of the model is broadly consistent with human performance; knowledge acquired in childhood may degrade over time through lack of use, but can be revived with modest additional experience.

We now consider the results for the Flat trajectory condition. This condition addresses the concern that the results of Simulation 1 might have derived from two complementary AoA effects: one due to high frequency of exposure early in training and one due to high frequency of exposure late in training. If this were correct, performance at the end of training in both the Early and Late conditions should be better than in the Flat condition, in which frequencies changed very little across epochs. This result was not observed. Figure 5 summarizes performance in the Flat condition and on the same items in the Early and Late conditions from Simulation 1.

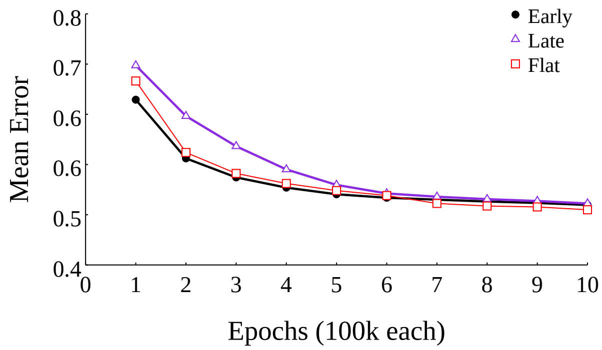

Figure 5. Performance in the flat condition (Simulation 2) compared to the same items in the early and late conditions in Simulation 1

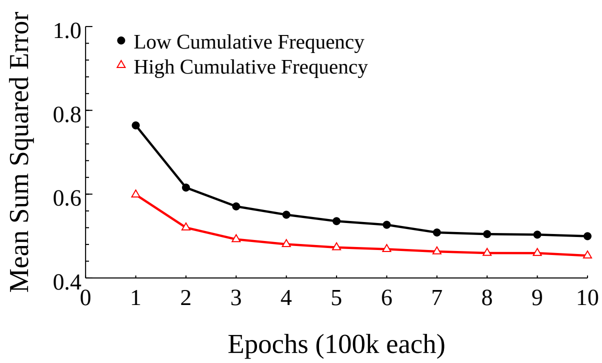

Figure 6. Performance on high and low cumulative frequency items within the flat condition

Results in the Flat condition closely resembled those obtained in the Early condition. Both conditions exhibited a small advantage early in training compared to the Late condition, but by the end of training all conditions converged on the same level of performance at the end of training. The mean SSE in the Flat condition was .48, compared to .48 and .49 in the Early and Late conditions respectively. No effect of frequency trajectory was observed, $F(1, 93) < 1$. The early advantage in the Flat condition reflects the fact that the items had a mean frequency of 20 presentations per 100,000, which was higher than in the Late condition over these epochs. However, the cumulative frequency of flat items (200) was not significantly different from the Early and Late items $F(1, 93) < 1$.

Data concerning the role of cumulative frequency are presented in Figure 6, which shows the sum squared error for the highest and lowest frequency 25 items. The mean cumulative frequencies for these subsets of these items differ: 544 for the highest frequency words and 60 for the lowest. Cumulative frequency has the expected effect on performance, which is better for high frequency words (.46) than low words (.55), $t(47) = 3.22, p < .005$. Note that these means are substantially lower than the means for the critical items in Simulation 2. This suggests that the failures to observe

AoA effects were not due to floor effects on the critical items.

Discussion

Results from the Early and Late conditions were consistent with Simulation 1. There was a larger difference between these conditions until well into training, which reflects the fact that the critical words have few neighbors and therefore performance does not benefit as much from training on other words. However, performance in the two training conditions again converged as the cumulative frequencies evened out. Thus the results of Simulation 1 generalize to stimuli that have less consistent spelling-sound mappings. Performance on words in the Flat condition converged to the same level as on these same words in the Early and Late conditions in Simulation 1, indicating that the results for the Early and Late conditions did not reflect two complementary types of facilitation. Finally, there was an effect of cumulative frequency in the Flat condition: at the end of training performance was better on the words with higher cumulative frequencies than lower.

These results suggest that whereas cumulative frequency has an impact on performance, frequency trajectory does not. The age of acquisition hypothesis tested in previous behavioral experiments was that there would be a residual effect of early word learning on skilled adult performance. However, although words in the Early condition were learned more rapidly than words in the Late condition, performance in the two conditions was nearly identical by the end of training.

Simulation 3

To this point the results suggest that when cumulative frequencies and stimulus properties are equated across conditions, there is little if any effect of frequency trajectory. What matters is how often a word is encountered, not the pattern of encounters over time. Here we consider another factor that may have contributed to these results: the fact that the training corpus consisted of words that exhibit systematic relationships between orthography and phonology. What the model learns about one word carries over to other words that share structure with it, reducing the effects of lexical frequency (Seidenberg & McClelland, 1989) and thus the effects of any frequency trajectory manipulation. These neighborhood effects were larger for the consistent words used in Simulation 1 than for the strange items used in Simulation 2; the consistent words were learned more rapidly and yielded better asymptotic performance than the strange words even though the trajectories and cumulative frequencies were very similar in the two cases. Although the strange words have fewer close neighbors, their orthographic-phonological correspondences are not arbitrary; a word such as BEIGE is not pronounced “glorp;” it overlaps with more distant neighbors such as BINGE, BARGE, WEIGH and many other words among the background stimuli. Thus the systematic aspects of the orthography $\rightarrow$ phonology mapping might have reduced trajectory effects even for the strange words.

Suggestive evidence is provided by simulations of age of acquisition effects presented by Ellis and Lambon Ralph (2000). Feedforward models were trained to produce a transformation of arbitrary bit vectors. In their training set, output vectors were generated by randomly changing 10% of the bits in the input vector. Ellis and Lambon Ralph (2000) observed strong age of acquisition effects, such that items that were introduced early had an advantage over late items, even when the later items were much higher in cumulative frequency. The nature of the stimuli meant that learning on any given trial carried little information relevant to other items. Under this condition, there was a residual advantage for mappings that became entrenched early in training. Ellis and Lambon Ralph (2000) provide a thorough discussion of why this entrenchment occurs. In essence, learning that occurs for early-trained items involves large weight changes that reduce the model's sensitivity to error signals generated by the presentation of later items. Smith et al. (2001) provide a similar analysis of the results of their simulation, which was also constructed so that what was learned on one trial did not carry over to other trials.

Together the results of Simulations 1-2 and the Ellis and Lambon Ralph (2000) and Smith et al. (2001) simulations suggest that the nature of the input-output mapping – specifically whether what is learned on one trial predicts anything about other trials – may be crucial to producing AoA effects. To investigate this hypothesis, we devised a training regime deliberately unlike the orthography $\rightarrow$ phonology translation in English. Items for the Early and Late trajectory conditions in Simulation 3 were constructed such that Early and Late items had minimal orthographic or phonological overlap. In addition, we did not include any background items; thus what the model learned depended solely on the properties of the critical stimuli. These conditions are more comparable to the ones studied by Ellis and Lambon Ralph and Smith et al. (2001) [5]

Methods

The training set consisted of 68 words. Two lists were created out of different inventories of letters and phonemes. One list included items such as COB, COG, COP, HOG, HOP, and TOG, whereas the other contained items such as BAD, BAN, BANE, PANE, PAN, and PAT. Some phonemes occurred in both lists (e.g., /p/), but in different positions in different lists (e.g., onset and coda). The model's phonological representation (Harm & Seidenberg, 1999) treats these as separate phonemes; thus what is learned about onset /p/ does not carry over to coda /p/. The simulation was run twice with lists assigned once to each trajectory condition (Early, Late). In contrast to Simulations 1-2, no other words were presented during training. Thus, the model could learn regularities among the items within a training condition, but these regularities did not extend to the items in the other list, and performance was not modulated by exposure to any non-critical items.

Due to the smaller size of the training set, the models in Simulations 3 and 4 used a scaled down architecture with 29 orthographic units, 40 hidden units and 10 cleanup units. The phonological layer was kept the same. Frequency trajectories for items in Simulations 3 and 4 were similar to those in Simulations 1 and 2. However, because no "background" items were present, the range between lowest (9 per 10000) and highest (290 per 10000) frequency words is more dramatic. This is because how frequently an item is presented depends on both its log-compressed frequency and the number of other items in the training set. In the previous simulations, nearly 3000 words were being trained, so that even items with very high frequencies were only seen, on average, about 40 times per 100,000 trials. In this simulation, only 68 items were trained, resulting in higher real frequencies, although the log compressed frequencies used to select items were the same. Also because of the smaller training set, fewer training trials were required: The model was trained for 10 epochs of 10,000 trials each, resulting in 100,000 training trials, as opposed to 1 million in Simulations 1 and 2. The mean cumulative frequency of Early words (1474) was not different from the cumulative frequency of Late words (1467), $t(67) < 1$.

[5] The simulations in this article were actually conducted before we were aware of the Ellis and Lambon Ralph (2000), Smith et al. (2001) or Monaghan and Ellis (in press) simulations.

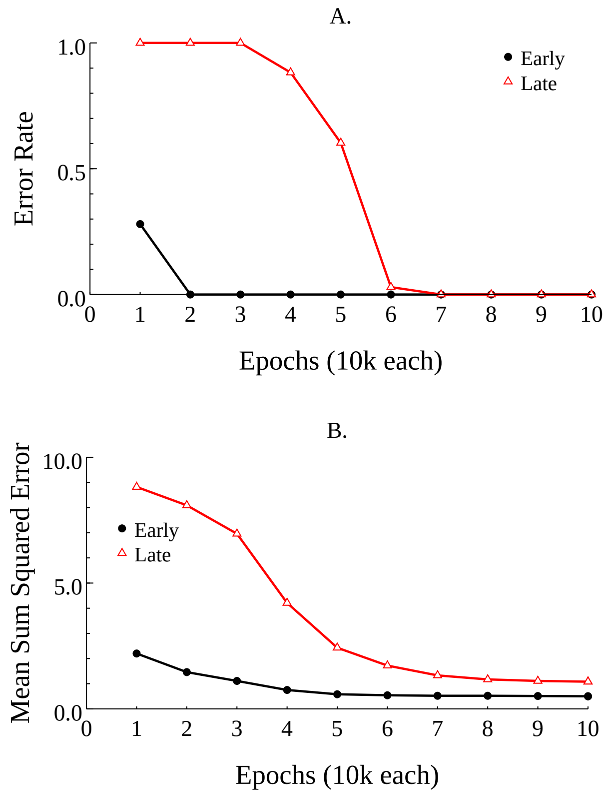

Figure 7. Performance over time for critical items in Simulation 3: A) error rate, B) sum squared error

Results and Discussion

Figure 7 presents the accuracy and mean SSE data over the course of training. By the end of training the model had learned to produce correct output for all words. Whereas all of the Early items were learned within the first 2 epochs, the Late items did not reach this level until much later. The mean number of trials to learn the Early words was 1.3 epochs vs 5.5 for the Late items, a highly reliable difference, $t(67) = 49.1$. Again, these numbers reflect the point in training as a function of all trials for all items. Because so many of the Early items were learned within the first epoch, the mean number of exposures before learning was computed by examining the model's performance at 1,000 trial intervals. By this measure, the mean number of exposures to a given item before it was learned was 242 for Early items and 270 for Late items. Note that this is different from Simulation 2, in which *fewer* actual exposures were required for the learning of the Late items. In this simulation, knowledge of the Early items seemed to impede rather than aid learning of the late items. The contrast provides a reminder of the extent to which learning spelling-sound correspondences normally depends on exposure to neighbors.

In contrast to previous simulations, there was a small but reliable advantage for words that were presented frequently early in training in Simulation 3, even after the cumulative frequencies in the Early and Late conditions converged. As shown in Figure 7B, there was an advantage for Early words that was maintained through 10 epochs of training. A t-test on the mean SSE at the end of training revealed that error was reliably greater for Late words (1.13), than Early words (.74), $t(67) = 10.08, p < .001$.

The critical difference between the simulations concerns the nature of the stimuli and thus the mapping between input and output codes. Simulations 1 and 2 used a large corpus of words that exhibit the regularities between spelling and sound characteristic of English orthography. These regularities modulate the effects of frequency of exposure to a given word, yielding no residual effect of frequency trajectory on skilled performance. This result obtains when other stimulus properties and cumulative frequencies are controlled.

In Simulation 3, the normal regularities in the mapping between spelling and sound were not maintained because we eliminated the background items and created nonoverlapping stimulus sets. What the model learned about one word in a training list carried over to other words on the same list, but not to words on the other list. Given this sharp dissociation between the stimulus characteristics of Early and Late words, there was an advantage for the early-trained items.

Simulation 4

Simulation 3 strongly suggests that the nature of the mapping between input and output determines whether frequency trajectory affects performance. However, this simulation differed from the earlier ones in a number of other ways (e.g., the number of units; size of the training corpus). We therefore ran a final simulation using the same procedures as in Simulation 3, but using stimuli which, like the ones in Simulations 1-2, contain overlapping orthographic and phonological patterns.

Methods

The same items from Simulation 3 were used, but rather than segregate items such that no letter or phoneme was repeated in the same position between lists, we organized the lists so that no letter or phoneme occurred on one list but not the other. For example HUB, HUG, LUCK, PAT, and MAD were on List 1, whereas HUCK, LOG, LUG, MATE, and PAD were on List 2. Cumulative frequency of Early (1474) and Late (1467) words was matched $t(67) = 1.12, p > .2$.

Results and Discussion

As in Simulations 2 and 3, Early items were learned quickly (1.7 epochs) whereas Late words required more training to be accurately named (3.7 epochs). This difference is reliable $t(67) = 9.8, p < .001$. This is reflected in the change in accuracy over time, shown in Figure 8A. Also note that accuracy on both Early and Late items reached 100% by the 6th epoch; thus, although frequency trajectory had the expected effect on AoA, it had no residual effect on accuracy. The model's ability to generalize from Early to late items meant that even though it took much *longer* in terms of training epochs for the Late items to be learned, they were produced correctly after many fewer trials per word: the mean number of exposures to produce correct output was 262 for Early items and 52 for Late. As shown in Figure 8B, sum squared error on the Late words decreased more slowly than for the Early words, but performance in the two conditions eventually converged. The SSE was not different between Early (1.13) and Late (1.13) items $t(67) < 1$ at the end of training. As in Simulations 1-2, there was no residual effect of frequency trajectory when cumulative frequencies were matched. Error declined much more rapidly for the Late words in this Simulation (Figure 9A) than in Simulation 3 (Figure 8A). This is because learning on the Early

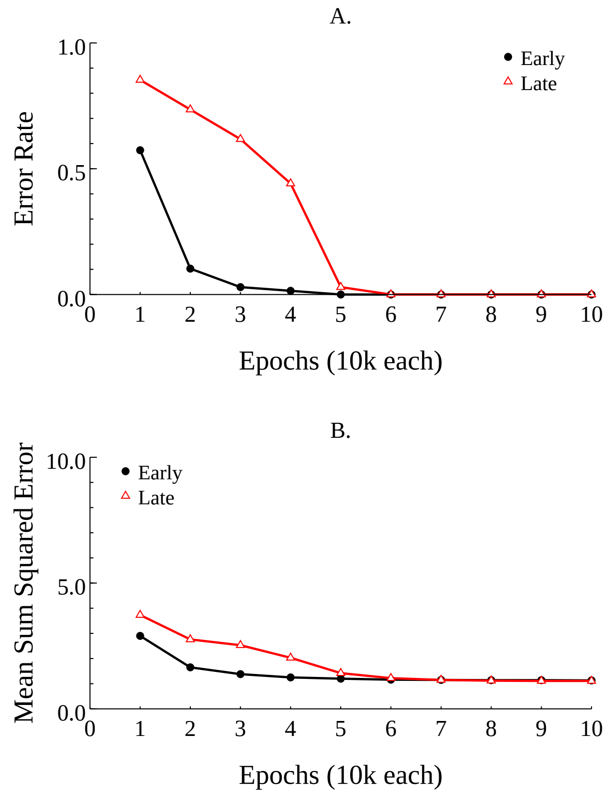

Figure 8. Performance over time for critical items in simulation 4: A) error rate, B) sum squared error

items transferred to performance on Late items, whereas in Simulation 3, learning on Early and Late items was independent.

Because this simulation was identical in every other respect to Simulation 3, the results indicate that the factor relevant to producing a frequency trajectory effect in Simulation 3 was the lack of overlap between Early and Late words.

General Discussion

Studies of age of acquisition effects have raised important questions about the effects of early experience on later learning. An effect of age of acquisition on skilled reading would call into question the results of many previous behavioral studies and models in which this factor was not investigated. The potential theoretical importance of this phenomenon as well as methodological and theoretical concerns led us to examine it further. Examination of the materials used in previous studies suggested that they did not provide strong evidence for an effect of age of acquisition independent of other measures of frequency with which AoA was confounded. The regression analyses provided evidence that age of acquisition ratings may account for a small amount of variance in skilled performance with other factors statistically controlled, but via the fact that they are correlated with how often words are used pre-adulthood. Thus there was no effect of AoA independent of cumulative frequency, as indexed by the WFG norms.

The results of Simulations 1 and 2 are consistent with these conclusions and provide evidence concerning the computational mechanisms that give rise to the behavioral phenomena. The simulations provide a strong test of the AoA hypothesis because the cumulative frequencies and frequency trajectories were known, and properties of early and late stimuli were equated exactly. The training corpus was a large, representative sample of monosyllabic words, which exhibit the statistical regularities characteristic of the orthography $\rightarrow$ phonology mapping in English. There was an initial advantage for words presented more frequently early in training, but no residual effect of early learning on skilled performance. This was true for both words with highly consistent spelling-sound correspondences (Simulation 1) and words with atypical spellings and pronunciations (Simulation 2). The advantage for early-trained words is washed out as the model picks up on the similarities that hold across words. This occurs more rapidly for words such as LAST whose component spelling patterns are pronounced consistently across many words than for strange words such as BEIGE which have fewer close neighbors. In both cases, however, early and late trained words converged to the same level of performance as the number of exposures evened out. This behavior can be traced to basic properties of connectionist models (Seidenberg & McClelland, 1989). Knowledge in these models is encoded in weights on connections among units, which reflect the cumulative effects of exposure to all words. Changes to the weights that occur when a word is trained also benefit words with which it overlaps. This leaves little room for early words to maintain an advantage, because the weights that support them also facilitate learning later-learned words.

Simulations 3 and 4 provided further evidence consistent with this analysis. In Simulation 3, we removed the overlap between early and late trained words and observed a reliable "age of acquisition" effect: there was an advantage for early-trained words that was maintained throughout the course of training. In this case, learning of the late items was impeded by the model's knowledge of the early-learned words. Finally, in Simulation 4, we reintroduced the overlap between early and late trained words and the age of acquisition effect was eliminated, further demonstrating that the critical factor that gave rise to the AoA effects in Simulation 3 was the lack of overlap among the early and late patterns.

In summary, both the behavioral data and the simulations are consistent with the conclusion that whereas there is an effect of cumulative frequency on reading performance, there is no independent effect of the age at which words are learned.

Conditions That Create Age of Acquisition Effects.

In the remainder of this article we consider other types of conditions and tasks for which age of acquisition effects are likely to be more prominent. Our Simulation 3 and the simulations previously reported by Ellis and Lambon Ralph (2000), Smith et al. (2001), and Monaghan and Ellis (in press) all suggest that age of acquisition effects will occur under some circumstances. Although these simulations differ in detail, they share an important property: given the nature of the stimuli and network architecture, what was learned about early-trained patterns did not carry over to later-trained patterns. Early-trained patterns became entrenched, yielding a persistent advantage over later-trained patterns. Our main point is that the conditions that give rise to these effects are not characteristic of reading an alphabetic orthography, but are potentially relevant to other tasks. To see this clearly, it is necessary to examine some details of the simulations.

The Ellis and Lambon Ralph (2000) simulations involved a simple feedforward network. The input and output layers each consisted of 100 units, and there were 50 hidden units. The input stimuli consisted of random bit patterns created by activating a random 20% of the units on the input layer. The model was trained to copy the input onto the output, but with 10% of the bit values changed (randomly determined in advance). Two aspects of the simulations underlie the strong age of acquisition effects that were observed. One has to do with the nature of the patterns that were trained and the other with the nature of the mapping between input and output.

The important property of the training patterns is that, unlike words in natural languages, they did not exhibit a rich internal structure. The statistical structure of the lexicon reflects the fact that there are constraints on the ordering of letters and phonemes and differences in the frequencies with which these elements occur and co-occur. Much of this structure ultimately derives from constraints imposed by speech perception and production; for example, certain sequences of phonemes are ruled out because they cannot easily be articulated; the relative frequencies of patterns are determined in part by ease of articulation; and so on. These constraints are also reflected in alphabetic writing systems because they are codes for representing speech. In contrast, the stimuli in the Ellis and Lambon Ralph simulation were constructed so that the probability that any given unit was on was independent of the probabilities for all other units. Under this condition, what is learned about one pattern does not carry information about other patterns. Using an architecture with a smaller number of hidden units than input or output units promotes the discovery of subregularities that hold across patterns (as occurs, e.g., with words). If these regularities do not exist, however, the model can only learn the task by memorizing individual patterns, even though the mapping is *prima facie* highly consistent. Under these conditions, early-trained patterns become entrenched: the large initial weight changes that favor these patterns are difficult for later-trained patterns to overcome.

The nature of the mapping between input and output codes also promoted pattern memorization in these simulations. The fact that the mapping between input and ouput involved random changes to 10% of the bits meant that the model could not generalize from early-trained patterns to later-trained ones accurately. The mapping between input and output codes contained a partial regularity (90% of the input bits mapped onto the corresponding output bit) but the inconsistent elements were random and therefore unlearnable except by memorization.

The Smith et al. (2001) simulation was similar in that the stimuli were random bit patterns that were not internally structured. Their model was also trained to copy the input to the output through a smaller number of hidden units, but without the random changes to 10% of the bits. Like Ellis and Lambon Ralph's model, Smith et al.'s performed the task by memorizing the training patterns, and again exhibited entrenchment of early-learned patterns.

The Monaghan and Ellis (in press) simulation also conforms to this analysis, although it differs from the other simulations in interesting ways. The simulation again involved a simple feedforward network. Unlike the simulations discussed above, the training patterns were designed to capture some aspects of lexical structure. The input and output layers were divided into three slots, analogous to a CVC syllablic structure. Within each slot there were ten bit patterns ("phonemes") that were repeated across stimuli in the training set. Thus there were constraints on which units could and could not be simultaneously activated; what was learned about one occurrence of a pattern over the whole set of input units could carry over to other patterns with which it overlapped – i.e., those containing the same "phonemes."

Monaghan and Ellis also manipulated the consistency of the mapping from input to output. In a behavioral experiment, they found that whereas words with inconsistent spelling-sound correspondences produced an age of acquisition effect, words with consistent correspondences did not. The stimuli in this study were discussed earlier; there is some evidence that the effect was due to frequency rather than age of acquisition. In the simulation of these effects, the consistency of the mapping from input layer to output was varied. On 80% of the trials, the model was trained to copy the input; on the other 20% the input the "consonants" were copied but the "vowel" was randomly assigned to one of the other 9 possible vowels. The consistent patterns did not produce an age of acquisition effect, whereas the inconsistent patterns did.

The results for the consistent condition are like those we observed in Simulations 1: no age of acquisition effect when the stimuli overlap in structure. The results for inconsistent patterns appear to conflict with the results of Simulation 2, in which we did not observe an age of acquisition effect for words with atypical ("inconsistent") spelling-sound correspondences. However, the differing results are traceable to properties of the stimuli. Our model was trained on a large set of words; the critical stimuli were a subset of "strange" words that contain atypical spelling-sound correspondences. The modeling indicates that these words nonetheless overlap sufficiently with other words in the corpus to wash out the initial advantage for early-trained items.

Monaghan and Ellis' inconsistent stimuli were wordlike patterns in which the "vowel" was randomly mapped onto other vowels for 20% of the items. Given the arbitrary nature of these mappings, the model could only perform the task by memorizing the patterns. As in other conditions in which patterns must be memorized, there was a strong age of acquisition effect. It is important to note that this degree of arbitrariness is not seen in English words, even strange ones. Although vowel graphemes in English map onto multiple phonemes, the range of possibilities is constrained. No vowel grapheme maps onto all possible vowels (Venezky, 1970); typically the irregular pronunciation is a small number of phonetic features away from the "regular" pronunciation. Thus HAVE is irregular, but /ae/, like /eI/ is a front, unrounded vowel, not a more distant vowel such as /oʊ/. This general pattern is also observed with other irregularly-pronounced vowels; for example, EA may be pronounced as in BEAD, BREAD and BREAK, all of which contain mid-to-high front, unrounded vowels (/i/, /ɛ/ and /eI/ respectively). A word like BEIGE is "strange" in the sense that it lacks immediate neighbors, but the EI $\rightarrow$ /eI/ mapping is supported by other words in the lexicon (WEIGH, EIGHT, HEIR). Finally, although vowel graphemes map onto multiple phonemes in English, the pronunciations are typically cued by surrounding letters. The regularities that exist over the units termed rimes (or "word-bodies") have been extensively studied, but there are partial regularities involving other parts of words as well (Kessler & Treiman, 2001). In Monaghan and Ellis's stimuli, the alternative pronunciations of vowels were assigned independently of context.

These examples illustrate only some aspects of the statistical structure of words in English. The important point is that the characterstics of the stimuli in the Monaghan and Ellis simulation were quite different, even though the simulation was intended to be relevant to consistency effects in English. Their stimuli produced large age of acquisition effects because they lacked the redundancy of English words.

In summary, all of the simulations of age of acquisition effects are consistent with the same conclusion: AoA effects depend on the nature of the mapping between codes, specifically whether what is learned about early-learned patterns carries over to later patterns. When the stimuli and task afford this type of learning, the network does not have to memorize individual patterns; it encodes regularities across patterns which allow the model to generalize, washing out the initial advantage for early-trained words. Simulations 1 and 2 provide the most direct evidence concerning such effects in reading, insofar as the model was trained on a large corpus of words exhibiting the spelling-sound mappings characteristic of English. When the stimuli and task do not afford this type of learning (the Ellis and Lambon Ralph (2000) and Smith et al. (2001) simulations, and Monaghan & Ellis's inconsistent condition), the network is forced to memorize patterns, yielding an advantage for early-trained ones. In this light it is interesting to consider our Simulation 3, in which the Early and Late items overlapped among themselves, but not across lists. In this case, the model could generalize from one Early item to another, and from one Late item to another, but the orthogonal nature of the lists made it such that the Late items as a group were learned suboptimally – the representations developed to support the Early items impeded acquisition of the Late items.

It should be noted that our simulations did not address all aspects of lexical processing and so cannot be taken as showing that such effects cannot occur. The simulations involved knowledge of orthographic $\rightarrow$ phonological correspondences and we have argued that they are consistent with behavioral studies of age of acquisition effects that used tasks, such as naming and lexical decision, in which this knowledge plays an important role. The simulations suggest that the age at which this knowledge is acquired has little impact on skilled performance. The original age of acquisition hypothesis (Brown & Watson, 1987; Morrison & Ellis, 1995) however, concerned the effect of the age at which words are acquired in spoken language, an aspect of lexical learning our simulations did not address. Acquiring a spoken word vocabulary involves learning mappings between phonology and semantics. Skilled reading often involves computations from orthography to phonology to semantics (see, e.g., Van Orden, Johnston, and Hale (1988), for behavioral evidence and Harm and Seidenberg (2001), for a computational model). Hence the age at which children learned phonology to semantics mappings could have a residual impact on the orthography $\rightarrow$ phonology $\rightarrow$ semantics computation. None of the simulations of age of acquisition effects, including our own, address this possibility.

This issue needs to be examined in future research. Two points should be noted, however. First, we have presented evidence that the results of existing behavioral studies can be explained in terms of the impact of lexical factors such as frequency, imageability and length on word reading. Thus, it is not clear if there is an age of acquisition effect to be explained further. Second, properties of the phonology $\rightarrow$ semantics mapping make it unlikely to be the source of effects of age of acquisition on reading. The mapping between these codes is largely arbitrary for monomorphemic words; words that overlap with the sound of the word CAT do not overlap with it in meaning. Thus what is learned about the phonology $\rightarrow$ semantics mapping for CAT does not carry information that facilitates learning the mapping for SAT or FAT. Given the computational analysis presented above, this might seem like a condition that would promote a strong age of acquisition effect in spoken language acquisition, which in turn could affect reading via the shared phonology $\rightarrow$ semantics pathway. However, other characteristcs of the phonology $\rightarrow$ semantics mapping need to be taken into account. First, the mapping between phonology and semantics is not entirely arbitrary; there are partial regularities among many monomorphemic words (e.g., correlations between the phonological characteristics of words and their grammatical class; Kelly, 1992); more importantly, inflectional and derivational morphemes make consistent (though quasiregular) contributions to the meanings of many words (Seidenberg & Gonnerman, 2000). Second, both phonology and semantics are themselves highly structured: the words of a language occupy restricted regions of the much larger space of possible phonological forms or meanings. All of these properties will facilitate the learning of mappings between phonology and semantics in many types of connectionist networks, reducing effects of the ages at which words are learned, as in the simulations presented above.

Which Types of Knowledge Yield Age of Acquisition Effects?

On our account, the key issue regarding age of acquisition effects concerns the nature of the stimuli and task being learned. The research discussed in this article, like the behavioral studies discussed above, focused on the use of information concerning orthographic-phonological correspondences in English. The analyses of previous studies, the theoretical analysis of the problem, and the results of the simulations all suggest that AoA effects are likely to be minimal in this domain. However, the modeling led to the identification of other conditions that give rise to age of acquisition effects. The question then is whether these conditions are characteristic of other types of human learning. This issue needs to be considered further using both behavioral and modeling approaches.

One obvious question is whether there are age of acquisition effects in reading nonalphabetic writing systems such as Chinese. Written Chinese exhibits less consistency in the mapping between written symbols (characters) and their pronunciations. Chinese words are usually taught as arbitrary associations between written words and meanings, a process requiring several years for the mastery of a few thousand characters. There may be a lasting advantage for early-learned words in Chinese because of the more arbitrary nature of the mapping. This unresolved empirical question needs to be addressed carefully. Many of the early-learned words are nonarbitrary in that they contain characters that provide partial cues to pronunciation. The same need to control for other correlated properties (e.g., frequency) will also arise. This is illustrated by recent studies of AoA effects in reading Kanji, the Chinese characters that are part of Japanese writing. Yamazaki, Ellis, Morrison, and Lambon Ralph (1997) reported data indicating an AoA effect on Kanji naming; however, further analyses by Yamada, Takashima, and Yamazaki (1998) suggest that other factors may be at work. They found that the ease with which naive students could learn the pronunciations of the characters in question was also a strong predictor of naming latency. Thus the effect seems to be due to stimulus factors other than age of acquisition.

AoA effects have been observed in several tasks other than reading. Many of these studies are also subject to the methodological concerns we have raised, but the findings are suggestive. One task that probably yields genuine AoA effects is learning the names associated with faces. Moore and Valentine (1998) studied this using faces rated for both subjective frequency and AoA. The earlier acquired faces were named more quickly than later acquired faces, with subjective frequency controlled. Moore and Valentine (1999) also found that AoA effects in face naming were stronger than those in name reading. Lewis (1999) found similar effects with faces from long-running soap operas, where more objective controls of the time at which individuals came in and out of public awareness were possible. Whereas Moore and Valentine attributed the effects to age of acquisition, Lewis interpreted them as effects of cumulative frequency. Although further research is needed, the effects are consistent with the theory presented here. Unlike words, face-name pairs provide a strong test of the AoA hypothesis, because the earlier acquired items do not vary predictably along other dimensions that make them easier to learn or recognize. Aside from partial phonological regularities in name gender (Cassidy, Kelly, & Sharoni, 1998) and various national/ethnic regularities (one rarely meets an Italian named Wong, for example), matching names to faces is essentially an arbitrary mapping in that what is learned early does not carry over to later items.

Recent studies of Dutch by Brysbaert, Lange, and Van Wijnendaele (2000) and Brysbaert, Van Wijnendaele, and De Deyne (2000) also yielded results consistent with our account. They found larger effects of AoA in Dutch on associate generation and semantic classification tasks than on word naming. Word associations have an arbitrary, learned component. The high association between pairs such as BREAD-BUTTER or HUSBAND-WIFE cannot be simply due to overlap in meaning because other pairs that overlap in meaning to a similar degree are not as highly associated (e.g., BREAD-CAKE; HUSBAND-MAN). Moreover, both associate generation and semantic classification tasks involve using knowledge about word meanings, not merely orthographic-phonological correspondences. The relationship between form (orthography or phonology) and meaning is much less systematic than the relationship between orthography and phonology; words that overlap in spelling tend to overlap in sound but not in meaning. Thus the age of acquisition effects observed in these tasks may be related to the use of this information. Further research is needed, however, to determine more definitively whether age of acquisition has an effect on the orthography $\rightarrow$ semantics or phonology $\rightarrow$ semantics mappings. Furthermore, any task that uses word meanings is open to difficulties establishing the chain of causality: Are early AoA words easy because they are early, or are they early because they are easy? This problem will require some ingenious methodological innovations before it can be solved.

Finally, consider the problem of learning a second language. It is well known that some aspects of language learning are easier for children than for adults (Johnson & Newport, 1989; Flege et al., 1999). The second language learning situation is one in which what is learned early in experience (the first language) is not highly predictive of what is to be learned in the later phase (the second language). Assuming that both languages make use of overlapping neural structures (see Perani et al., 1998, for an interesting discussion) it follows that second language learning should be disadvantaged. On this view, so-called "sensitive period" effects are actually extreme cases of AoA effects – failures to learn in later life which reflect the entrenchment of early-learned patterns – and not maturational changes in the neural substrate supporting language acquisition, as has been classically presumed (Lenneberg, 1967; Neville & Bavelier, 2000). Further progress in understanding how early experience interacts with learning later in life will be facilitated by examining tasks in which such effects are likely to be most powerful, and by further exploring the computational mechanisms underlying these tasks.

Conclusions

The purpose of our research was to examine age of acquisition effects on skilled reading, a topic with potentially broad theoretical implications that has been the focus of considerable research. Ironically, the main conclusion to be drawn from our research is that age of acquisition effects are likely to occur, but for tasks other than reading an alphabetic orthography. Age of acquisition effects reflect a loss of plasticity associated with success in mastering a task, a phenomenon that occurs in many types of learning and species. The zebra finch's success in acquiring its characteristic song imposes significant constraints on its ability to acquire additional vocal behavior (Doupe & Kuhl, 1999). Similarly, the child's success in acquiring the phonological inventory or syntax of a language may constrain its ability to learn other languages (Johnson & Newport, 1989; Werker & Tees, 1984). Issues concerning the nature and limits of plasticity in different domains and their neural and computational bases are central ones in cognitive neuroscience. Connectionist models provide a computational framework for understanding plasticity in terms of the nature of the material to be learned, and how what is to be learned is affected by what has already been learned. The entrenchment phenomenon discussed above is one outcome that occurs in such networks and we have taken a step toward specifying the conditions that give rise to it. Under other conditions, other outcomes are observed; in the reading case studied here, later learning is facilitated by prior knowledge rather than restricted by it. In the catastrophic interference case (McCloskey & Cohen, 1989), later success in learning results in forgetting of earlier material. Gaining a deeper understanding of the principles that govern the entire set of outcomes, and how they relate to the various tasks that humans perform, is an important goal for future research.

References

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CD-ROM).* (Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA)

Balota, D. (1994). Visual word recognition:The journey from features to meaning. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (p. 303-356). San Diego, CA: Academic Press.

Balota, D., & Chumbley, J. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 340-357.

Balota, D., Pilotti, M., & Cortese, M. J. (2001). Item-level analyses of lexical decision performance: Results from a mega-study. *Memory & Cognition*. (manuscript submitted for publication)

Balota, D. A., & Abrams, R. A. (1995). Mental chronometry - beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 1289-1302.

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford: Clarendon Press.

Brown, G. D. A., & 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 & Cognition*, *15*, 208-216.

Brysbaert, M., Lange, M., & Van Wijnendaele, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology*, *12*, 65-85.

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, *104*, 215-226.

Butler, B., & Hains, S. (1979). Individual differences in word recognition latency. *Memory & Cognition*, *7*, 68-76.

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, *25*, 85-95.

Cassidy, K. W., Kelly, M. H., & Sharoni, L. (1998). Inferring gender from name phonology. *Journal of Experimental Psychology: General*, *128*, 362-381.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589-608.

Coltheart, M., Davelaar, E., Jonasson, K., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention & Performance VI* (p. 535-555). Hillsdale, NJ: Erlbaum.

Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology*, *79*, 1-12.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, *22*, 567-631.

Ellis, A., & Morrison, C. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 515-523.

Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 1103-1123.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*, 78-104.

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

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. (1999a). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. *Memory & Cognition*, *27*, 592-602.

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

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*, 256-281.

Gilhooly, K. J., & Gilhooly, M. L. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology*, *71*, 105-110.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods and Instruments*, *12*, 395-427.

Harm, M. W. (1998). *Division of labor in a computational model of visual word recognition.* Unpublished doctoral dissertation, University of Southern California, Los Angeles, CA.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading, and dyslexia: Insights from connectionist models. *Psychological Review*, *163*, 491-528.

Harm, M. W., & Seidenberg, M. S. (2001). *Division of labor in a mulitcomponent model of visual word recognition.* (manuscript submitted for publication)

Hetherington, P., & Seidenberg, M. S. (1989). Is there "catastrophic interference" in connectionist networks? In *Proceedings of the 11th annual conference of the cognitive science society* (p. 26-33). Hillsdale, NJ: Erlbaum.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, *98*(1), 74-95.

Hirsh, K. W., & Ellis, A. W. (1994). Age of acquisition and lexical processing in aphasia: A case study. *Cognitive Neuropsychology*, *11*, 435-458.

Hodgson, C., & Ellis, A. W. (1998). Last in, first to go: Age of acquisition and naming in the eldery. *Brain and Language*, *64*, 146-163.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60-99.

Kawamoto, A. H., Kello, C. T., Jones, R. J., & Bame, K. (1998). Initial phoneme versus whole word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 862-885.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, *99*, 349-364.

Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in english monosyllables. *Journal of Memory and Language*, *24*, 592-617.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Lambon Ralph, M. A., Graham, K. S., Ellis, A. W., & Hodges, J. R. (1998). Naming in semantic dementia – what matters? *Neuropsychologia*, *36*, 775-784.

Lenneberg, E. H. (1967). *Biological foundations of language.* New York: Wiley.

Lewis, M. B. (1999). Are age-of-acquisition effects cumulative-frequency effects in disguise? A reply to Moore, Valentine and Turner (1999). *Cognition*, *72*, 311-316.

Lyons, A., Teer, P., & Rubenstein, H. (1978). Age-at-acquisition and word recognition. *Journal of Psycholinguistic Research*, *7*, 179-187.

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, *21*, 339-366.

Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, *19*, 313-330.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, *385*, 813-815.

McCandliss, B. D., Posner, M. I., & Givon, T. (1997). Brain plasticity in learning visual words. *Cognitive Psychology*, *33*, 88-110.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 23, p. 109-164). New York, NY: Academic Press.

Monaghan, J., & Ellis, A. W. (in press). What exactly interacts with spelling-sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory and Cognition*.

Moore, V., & Valentine, T. (1998). The effect of age of acquisition on speed and accuracy of naming famous faces. *The Quarterly Journal of Experimental Psychology*, *51A*, 485-513.

Moore, V., & Valentine, T. (1999). The effects of age of acquisition in processing famous faces : Exploring the locus and proposing a mechanism. In *Proceedings of the twenty-first annual conference of the cognitive science society* (p. 416-421). New Jersey: Erlbaum.

Morrison, C. A., & Ellis, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology*, *91*, 167-180.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 116-153.

Morrison, C. M., Ellis, A. W., & Chappell, T. D. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *50A*, 528-559.

Munro, P. W. (1986). State-dependent factors influences neural plasticity: A partial account of the critical period. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (p. 471-502). Cambridge, MA: MIT Press.

Neville, H. J., & Bavelier, D. (2000). Specificity and plasticity in neurocognitive development in humans. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (p. 83-98). Cambridge, MA: MIT press.

Perani, D., Paulesu, E., Galles, N., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S., Fazio, F., & Mehler, J. (1998). The bilingual brain - Proficiency and age of acquisition of the second language. *Brain*, *121*, 1841-1852.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56-115.

Quartz, S., & Sejnowski, T. J. (1997). The neural basis of cognitive development: A constructivist manifesto. *Brain and Behavioral Sciences*, *20*, 537-596.

Seidenberg, M. S. (1995). Visual word recognition: An overview. In P. Eimas & J. L. Miller (Eds.), *Handbook of perception and cognition: Language.* New York: Academic Press.

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, *4*, 353-361.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523-568.

Seidenberg, M. S., & Waters, G. S. (1989). Word recognition and naming: A mega study. *Bulletin of the Psychonomic Society*, *27*, 489.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, *23*, 383-404.

Smith, M. A., Cottrell, G. W., & Anderson, K. L. (2001). The early word catches the weights. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems 13* (p. 52-58). Cambridge, MA: MIT Press.

Spieler, D. H., & Balota, D. A. (1997). Connectionist models of word naming: An examination of item level performance. *Psychological Science*, *8*, 411-416.

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.

Turner, J. E., Valentine, T., & Ellis, A. W. (1998). Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. *Memory & Cognition*, *26*, 1282-1291.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from the spelling to sound to meaning. *Journal of Experimental Psychology: Memory, Language and Cognition*, *14*, 371-386.

Venezky, R. L. (1970). *The structure of English orthography.* The Hague: Mouton.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*(1), 49-63.

Yamada, J., Takashima, H., & Yamazaki, H. (1998). Effect of ease-of-acquisition on naming latency for Japanese Kanji: A reanalysis of Yamazaki et al.'s (1997) data. *Psychological Reports*, *83*, 991-1002.

Yamazaki, M., Ellis, A. W., Morrison, C. M., & Lambon Ralph, M. A. (1997). Two age of acquisition effects in the reading of Japanese Kanji. *British Journal of Psychology*, *88*, 407-411.

Zeno, S. (Ed.). (1995). *The educator's word frequency guide.* Brewster, NJ: Touchstone Applied Science Associates.

Appendix – Stimuli for All Simulations

Simulation 1

List 1	List 2	List 1	List 2
bail	beast	hatch	hoard
bay	beet	hound	hunt
belt	bill	maze	main
bench	bin	moist	match
bent	bit	mope	mist
blimp	bleat	pare	par
board	bound	pinch	pipe
broil	brag	pool	purse
cap	cab	quit	quench
car	care	seem	sift
cheat	cart	serve	sight
clip	chimp	skirt	skit
cog	clam	slam	slip
core	coat	street	stand
crass	cool	stuck	stick
curse	crab	stunt	stray
face	fail	swift	swerve
feast	fat	tab	tag
fill	felt	tart	tap
fine	fin	tight	teem
fist	flirt	tin	tent
flit	flog	toil	tore
float	foist	train	trope
grab	grace	trick	truck
grin	grass	twill	twist
grist	grill	vat	vine
hand	haze	wipe	winch

Simulation 2

List 1	List 2
ache	aisle
beige	bough
broad	brooch
caste	chaise
chic	choir
clique	coup
draught	ewe
friend	gaffe
gauge	ghoul
hearth	heir
hymn	
month	myrrh
pear	phlegm
pint	plaid
plaque	psalm
queue	realm
rheum	rogue
scheme	scourge
sew	shoe
sieve	ski
sponge	sword
touch	vague
valse	veldt
womb	young

Simulation 3

List 1	List 2
bad	cob
ban	cog
bane	cop
bat	cub
bate	flog
bid	flop
bide	hog
bin	hop
bit	hub
bite	huck
fad	hug
fade	log
fan	luck
fat	lug
fate	plop
fin	pluck
fine	plug
fit	roll
mad	rug
made	slob
man	slop
mane	sop
mat	stop
mate	stub
mid	stuck
mit	sub
mite	suck
pad	tog
pan	toll
pane	top
pat	troll
pin	truck
pine	tub
pit	tuck

Simulation 4

List 1	List 2
bad	ban
bane	bat
bate	bid
bide	bin
bit	bite
cob	cog
cop	cub
fad	fade
fan	fat
fate	fin
fine	fit
flog	flop
hog	hop
hub	huck
hug	log
luck	lug
mad	made
man	mane
mat	mate
mid	mit
mite	pad
pan	pane
pat	pin
pine	pit
plop	pluck
plug	roll
rug	slob
slop	sop
stop	stub
stuck	sub
suck	tog
toll	top
troll	truck
tub	tuck