ELSEVIER

Acta Psychologica 115 (2004) 43–67

acta psychologica

www.elsevier.com/locate/actpsy

Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation

Mandy Ghyselinck [a,*], Michael B. Lewis [b], Marc Brysbaert [c]

[a] *Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Gent, Belgium*
[b] *Cardiff University, P.O. Box 901, Cardiff CF1 3YG, UK*
[c] *Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK*

Received 20 July 2003; received in revised form 24 October 2003; accepted 17 November 2003

Abstract

Early-acquired words are processed faster than late-acquired words. This is a well-accepted effect within the word recognition literature. Different explanations have been proposed, either localizing the effect of age of acquisition (AoA) in a particular substage of word processing or seeing it as the result of the way in which information is stored and accessed in the brain in general. The cumulative-frequency hypothesis is an example of the latter type of explanation: It states that the total number of times a system has come across a particular stimulus will determine the speed with which the stimulus can be recognized. The present multi-task investigation provides a critical test of the different explanations. Results show that in a variety of word processing tasks the effects of frequency and AoA are highly correlated, and that the impact of AoA is consistently higher than would be expected on the basis of the cumulative-frequency hypothesis. The findings are interpreted as evidence for recent demonstrations of a loss of plasticity in neural networks due to training and/or for mathematical models that describe the growth of the lexico-semantic network as the attachment of new nodes to existing nodes.

PsycINFO classification: 2343
Keywords: Age of acquisition; Cumulative-frequency hypothesis; Lexical decisions; Naming; Perceptual identification

* Corresponding author. Tel.: +32-9-264-64-06; fax: +32-9-264-64-96.
E-mail address: mandy.ghyselinck@ugent.be (M. Ghyselinck).

0001-6918/$ - see front matter
doi:10.1016/j.actpsy.2003.11.002

1. Introduction

What makes some words faster to be recognized, named, or categorized? Why are certain words more easily responded to in decision tasks, skipped more often in reading, or pronounced faster when participants are required to name them as rapidly as possible? Throughout the history of psycholinguistic research, a number of variables have been proposed to answer these questions. Among them are word length, imageability, neighborhood size, neighborhood frequency, family size, concreteness, complexity, different measures of frequency of occurrence, familiarity, and age of acquisition. Of these variables, the most widely accepted is word frequency, as estimated on the basis of the number of times a particular word is encountered in a representative sample of texts or speech. This variable has been the subject of many studies and has become accepted as the most important variable in visual word recognition theories: Words that are often encountered in texts are recognized faster and/or more easily than words that are rarely seen.

In recent years, however, the primacy of word frequency has been challenged by the demonstration that the frequency effect is largely attenuated (and sometimes eliminated) when stimuli are controlled for the age at which they were learned. Because many high-frequency words are acquired relatively early in life and many low-frequency words relatively late in life, there is a natural confound between frequency of occurrence and age of acquisition (AoA) in unselected samples of stimuli. Consequently, word frequency effects found with uncontrolled samples could actually be due to AoA. Although the discussion has origins back to the 1970s (e.g., Carroll & White, 1973) and enjoyed some popularity in British psychology in the 1980s (e.g., Brown & Watson, 1987; Gilhooly & Watson, 1981), it only came to the foreground of research after Morrison and Ellis (1995) published an article in which they claimed that all frequency effects in word naming could be interpreted as AoA effects in disguise. In particular, Morrison and Ellis created a pair of word lists that differed in frequency and were matched on AoA and another pair of lists that differed in AoA and were matched on word frequency. Word naming times only differed for the lists in which AoA was varied, not for those in which frequency was manipulated. The article of Morrison and Ellis (1995) was followed by a whole series of studies designed to further investigate the claim that word frequency no longer had an effect in word naming if AoA was controlled for. In general, these studies confirmed that AoA was a significant variable in word naming, but also showed that Morrison and Ellis had underestimated the frequency effect, because combined effects of frequency and AoA were obtained (Brysbaert, 1996; Brysbaert, Lange, & Van Wijnendaele, 2000; Gerhand & Barry, 1998, 1999a; Morrison & Ellis, 2000; but see Barry, Hirsch, Johnston, & Williams, 2001).

AoA also proved to be a significant variable in other word processing paradigms, such as lexical decision (Brysbaert, Lange, et al., 2000; Gerhand & Barry, 1999b; Morrison & Ellis, 1995, 2000), word-associate generation and semantic categorization (Brysbaert, Van Wijnendaele, & De Deyne, 2000), picture naming (Barry, Morrison, & Ellis, 1997; Bonin, Chalard, Méot, & Fayol, 2002; Ellis & Morrison, 1998), speeded word naming (Gerhand & Barry, 1999a), and auditory lexical decision

(Turner, Valentine, & Ellis, 1998). In addition, effects of AoA were reported for face naming (Moore & Valentine, 1998) and face categorization (Lewis, 1999a; Moore & Valentine, 1999).

2. A genuine or confounded AoA effect?

Confronted with the above empirical evidence for AoA effects, the first reaction of many researchers was to ask for evidence that the effect was not caused either by the way in which AoA was measured or by a confounded variable that was not taken into account. As for the former, researchers raised objections against the fact that most AoA measures were based on student ratings. Typically, a group of undergraduates was asked to indicate on a 7-point scale when they thought they had learned different words, ranging from the age below 3 years to above 12 years (e.g., Gilhooly & Logie, 1980). It is not inconceivable that these ratings take other word characteristics into account, including the frequency of occurrence in daily life. To counter this criticism, adherents of the AoA view tried to validate their measures by correlating them with other estimates of AoA, such as the age at which words were supposed to be known by children in school (e.g., Ghyselinck, De Moor, & Brysbaert, 2000; Gilhooly & Gilhooly, 1980) and the proportion of children of different ages that were able to indicate the meaning of the words (e.g., Bonin, 2001; De Moor, Ghyselinck, & Brysbaert, 2000; Morrison, Chappell, & Ellis, 1997). Morrison et al. (1997), for instance, derived "objective AoA norms" by presenting a set of more than 200 object pictures to 14 groups of children aged 2 years 6 months to 10 years 11 months, plus an adult group, and calculated the age at which at least 75% of the children could name the object with or without a cue as to the initial sound of the word. A similar procedure was followed by Bonin in French. Both groups found a high correlation between the objective norms and the subjective ratings, and in subsequent studies reported that the objective AoA norms accounted for as much of the variance as the subjective ratings in picture naming (Bonin, 2001; Ellis & Morrison, 1998), word naming and lexical decision (Morrison & Ellis, 2000). Bonin even found that the objective AoA norms accounted for more of the variance in picture naming times. So, there is very little evidence indicating that the AoA effect would be a spurious finding due to the use of retrospective estimates.

Another criticism against the AoA effect has been that it is the confound of another variable. In its strongest version, this criticism says that a genuine effect of AoA can only be demonstrated if the variability due to all other possible word characteristics has been partialled out. These characteristics include word length in letters and phonemes, consistency and complexity of the letter–sound correspondences, bigram and trigram frequencies, neighborhood size and neighborhood frequency, family size, concept familiarity, imageability, and concreteness, and many other variables. Because of the multitude of word variables, which are all interrelated, it is virtually impossible to refute this criticism on empirical grounds. There will always be a combination of variables that accounts for most of the variance that could be attributed to AoA. To counter this criticism, researchers can only show that an explanation in terms of AoA forms the theoretically most coherent and parsimonious account of different empirical findings (for instance, below we will show that an AoA effect is actually required in many models that incorporate a frequency effect). They also must ascertain that the AoA effect is not a side-effect of some of the most obvious alternative candidates. So, quite some efforts have been invested, for example, to ensure that the AoA effect cannot be attributed to word imageability.

Word imageability (i.e., the ease with which a participant can create an image of the concept referred to by the word) has been proposed by several authors as a significant variable in word naming and lexical decision, because imageability is thought to be one of the most important variables in the organization of the semantic system and because the semantic system is believed to be involved in word naming and lexical decision (e.g., Balota, 1994; Baluch & Besner, 2001; Strain, Patterson, & Seidenberg, 1995; Zevin & Seidenberg, 2002). However, all studies that simultaneously manipulated AoA and imageability consistently reported a robust AoA effect when stimulus materials were controlled for imageability but very small or no imageability effects when stimuli were matched on AoA (Brysbaert, Lange, et al., 2000; Brysbaert et al., 2000; Coltheart, Laxon, & Keating, 1988; Morrison & Ellis, 2000). Brysbaert, Lange, et al. (2000) and Brysbaert et al. (2000), for example, created six lists of 24 Dutch words. The first pair of lists differed in frequency and was matched on AoA and imageability. The second pair differed in AoA and was matched on frequency and imageability. The third pair differed in imageability and was matched on frequency and AoA. Brysbaert, Lange, et al. (2000) reported significant frequency effects of 12 and 85 ms in naming and lexical decision respectively, together with significant AoA effects of 11 and 52 ms, but null-effects of imageability (1 and 0 ms respectively). The latter was unlikely to be due to the small range of imageability values used, as this variable yielded a healthy 279 ms effect in a semantic task (word-associate generation), together with a significant effect of AoA (279 ms) and an inverse effect of frequency (−218 ms; Brysbaert et al., 2000).

3. Different explanations of the AoA effect

Given that the AoA effect cannot be dismissed easily as an artifact due to the way it is measured or to some uncontrolled third variable that correlates with the age of acquisition, theorists have started to look for reasons why one would expect an effect of this variable. Probably the simplest model that predicts a combined effect of frequency and AoA is a model that sees the total number of encounters with a stimulus as the determining factor of processing speed. In such a cumulative-frequency model, both the number of times a stimulus is encountered per time unit (estimated by word frequency) and the total time the stimulus is known to the participant (estimated by word AoA) determine the ease with which a stimulus is recognized. In contrast, a model that predicts an effect of frequency without an accompanying effect of AoA either assumes that the word frequency measure is a good index of the total number of times an individual has come across a stimulus in his or her lifetime (i.e., the cumulative frequency) or that only the encounters in the most recent time period are important (so that differences in number of years-known do not have an effect). Although the cumulative-frequency hypothesis seems the most straightforward interpretation of combined frequency and AoA effects, it has only recently been proposed as a viable alternative (Lewis, 1999a, 1999b; Lewis, Gerhand, & Ellis, 2001; Zevin & Seidenberg, 2002) and at present is certainly not the most widely cited explanation of the observed AoA effects.

The first reason why researchers have rejected the cumulative-frequency hypothesis is that in most studies additive effects of frequency and AoA have been reported (e.g., Gerhand & Barry, 1998; see also the experiments below). Following Sternberg's (1969) stage theory, many authors interpreted this finding as evidence for a different origin of both effects. Otherwise, they argued, the frequency effect should be larger for early-acquired words than for late-acquired words, leading to a significant interaction between both variables. However, Lewis and colleagues (Lewis, 1999a; Lewis et al., 2001) showed that this argument is only valid if one accepts a linear relation between cumulative frequency and word processing time. If, as is usually done, a compressed function of frequency is adhered to, then additive effects of word frequency and word AoA ought to be expected. For instance, most researchers assume that the logarithm of word frequency better describes the frequency effect than the raw frequency data. Applied to the cumulative-frequency hypothesis, this means that

$$\begin{aligned} \mathrm{RT} &= a + b * \log(\text{cumulative frequency}) && \text{(1a)} \\ &= a + b * \log(\text{frequency} * \text{number of years-known}) && \text{(1b)} \\ &= a + b * \log(\text{frequency}) + b * \log(\text{number of years-known}) && \text{(1c)} \end{aligned}$$

(a and b are free parameters of the model).

A similar additivity is predicted if the compressive function is not a logarithmic function but a power function (e.g., Logan, 1988; Newell & Rosenbloom, 1981). For then

$$\begin{aligned} \mathrm{RT} &= a + b * (\text{cumulative frequency})^n && \text{(2a)} \\ \mathrm{Log}(\mathrm{RT} - a) &= \log(b * (\text{cumulative frequency})^n) && \text{(2b)} \\ &= \log(b) + n * \log(\text{cumulative frequency}) && \text{(2c)} \\ &= \log(b) + n * \log(\text{frequency}) && \text{(2d)} \\ &\quad + n * \log(\text{number of years-known}) && \text{(2e)} \end{aligned}$$

Lewis et al. (2001) showed that many of the previously collected English data (Balota, Cortese, & Pilotti, 1999; Carroll & White, 1973; Gerhand & Barry, 1998; Spieler & Balota, 1997) could well be described by the power function.

A second argument against the cumulative-frequency hypothesis has been that it predicts a decreasing impact of AoA as the participants grow older (this is true whether the frequency function is a linear function or a compressed function). Morrison, Hirsh, Chappell, and Ellis (2002), however, found no evidence for this prediction.

If anything, the AoA effect tended to be larger for the older age group than for the younger age group. Unfortunately, one has to be careful with this type of evidence if it is based on cross-sectional research (i.e., comparing the performance of younger and older participants at the same moment in time). Research on the relationship between age and intelligence has shown us how easily cohort effects can influence the findings (e.g., Schaie, 1990).

A third argument against the cumulative-frequency hypothesis has been that the AoA effect is most clearly present in tasks that require verbal output, whereas the clearest frequency effects are reported for tasks that capitalize on the input processes. For instance, Gilhooly and Logie (1981) failed to find an effect of AoA in a perceptual identification task with tachistoscopically presented words. Similarly, Morrison, Ellis, and Quinlan (1992) observed a strong AoA effect in object naming, but not in an object categorization task in which participants were to decide whether a picture represented a naturally occurring object or a man-made object. Findings like these (see also Barry et al., 2001) have led to the hypothesis that the origin of the AoA effect is situated either in the phonological output lexicon or in the connections between the semantic system and the lexical output phonology, whereas the frequency effect would have its origin in the input stages (e.g., Barry et al., 2001; Brown & Watson, 1987; Gerhand & Barry, 1998; Gilhooly & Watson, 1981; Morrison & Ellis, 1995).

The phonological output hypothesis, however, rests uneasy with repeated demonstrations of AoA effects in tasks that do not require phonological information. Although it can be defended that phonological information is used in lexical decision, explaining the AoA effect for this task (e.g., Gerhand & Barry, 1999b; Morrison & Ellis, 1995), it is more difficult to sustain such a position in the case of noun/first-name decisions (Brysbaert et al., 2000) and face recognition (Lewis, 1999a; Moore & Valentine, 1999). In addition, evidence has been reported suggesting that AoA may have an effect on the input processes of word recognition as well. Yamazaki, Ellis, Morrison, and Lambon Ralph (1997) showed single Kanji-words to participants and asked them to name the words as quickly as possible. Regression analyses showed that both the age at which the words had entered the spoken vocabulary and the age at which the written characters had been learned were predictors of reading speed. Brysbaert, Lange, et al. (2000, Experiment 2) obtained an AoA effect in a masked priming paradigm using orthographic neighbors. They asked participants to perform a lexical-decision task on words that were preceded by tachistoscopically presented prime words. Targets were late-acquired words, primes were either early-acquired neighbors (i.e., words that differed by one letter only; e.g., *fear-feat*) or unrelated words (*book-feat*). Participants took more time to accept a late-acquired target word after an early-acquired prime word than after an early-acquired unrelated word, just like they took more time to accept a low-frequency word after a high-frequency prime than after a high-frequency unrelated word. In general, the masked priming procedure is considered as a task that taps into the input processes (e.g., Segui & Grainger, 1990).

Brysbaert, Lange, et al. (2000) and Brysbaert et al. (2000) further suggested that there were theoretical grounds to expect an effect of AoA in the organization of the semantic system. As the semantic system is a highly structured network that is not acquired at once and in which the meaning of late-acquired concepts is often defined in terms of previously acquired concepts, it seems likely that the order in which concepts have been learned is not completely overruled by subsequent differences in frequency of occurrence (a point previously made by van Loon-Vervoorn (1989) as well). Recently, Steyvers and Tenenbaum (submitted for publication) have articulated a more detailed outline of this suggestion. They started from the observation that many growing networks possess a small-world structure (Milgram, 1967; Watts & Strogatz, 1998). This structure consists of a relatively small number of well-connected nodes that serve as hubs and that are the origin of clusters within the general architecture. Such a network has a number of very detailed mathematical characteristics which Steyvers and Tenenbaum were able to confirm for the semantic system on the basis of different data sets (e.g., the free association norms collected by Nelson, McEvoy, & Schreiber (1999)). In addition, it is known that the small-world structure is likely to be the result of a growth process that is governed by the preferential attachment principle. This principle says that new nodes are preferentially attached to existing nodes that already have a lot of connections. Because of that rich-gets-richer principle, there is a causal relationship between the history of a network's growth and its ultimate pattern of connectivity. On the whole, older nodes will possess more connections than younger nodes, even if variation in frequency (utility of the nodes) is allowed to modulate the probability of connecting new nodes to particular existing nodes. On the basis of this model, Steyvers and Tenenbaum predicted effects of both AoA and frequency in semantic tasks (and in lexical tasks if one assumes that word form and meaning are closely interconnected, as the authors seem to be doing).

Although Steyver and Tenenbaum's growing network model predicts combined effects of AoA and frequency in word processing tasks just like the cumulative-frequency hypothesis, there is one major difference. According to the cumulative-frequency hypothesis, frequency and years-known must have the same weight in the equation that predicts the response times (see the parameters b in Eq. (1c) and n in (2e)). The growing network model does not necessarily make this claim, as the importance of AoA relative to frequency depends on the weight given to the utility parameter for determining the likelihood of connecting new nodes to particular existing nodes.

Different weights of AoA and frequency are also predicted by a last class of models that have been proposed to account for the AoA effect. These models come from the connectionist tradition (Ellis & Lambon Ralph, 2000; Smith, Cottrell, & Anderson, 2001) and have shown that in a three-layer network activation patterns that are introduced first have a long-lasting advantage over later introduced patterns. This might be surprising at first, because connectionist models are typically associated with a phenomenon called "catastrophic interference" whereby newly trained patterns overwrite and eradicate old patterns. Catastrophic interference (new information is superior to old information) is exactly the opposite of the AoA effect (old information is superior to new information), and this idea was indeed originally used by Morrison and Ellis (1995) to argue that the AoA effect questioned the utility of connectionist models to understand visual word recognition. However, Ellis and Lambon Ralph (2000) subsequently showed that catastrophic interference is limited to those conditions under which the old information is *replaced* by new information. When the old information keeps on being presented to the model, together with the new information, in an interleaved way, connectionist models show a genuine AoA effect. Ellis and Lambon Ralph (2000) argued that the latter situation, with interleaved learning, is what happens when a child acquires a vocabulary. 'Old' words do not cease to be used when new words are learned. Instead, they keep on being used together with the later acquired words in an interleaved manner. It is therefore perfectly possible to apply connectionist principles to simulate and explain the AoA effect.

Connectionist models with distributed representations learn to recognize input patterns by changing the weights between the units in the network until the pattern of output produced by the network matches the desired output to a sufficient degree. Because of the way in which the weights are adjusted (usually according to the back-propagation algorithm), changes in weight are larger when the activation strength between two units is in the middle range (usually around 0.5) than when the activation strength is already close to one of the extremes (either 0.0 or 1.0). This implies that input patterns that are trained first (when the activations are all still in the middle range) are likely to induce larger weight shifts than patterns that are trained later when the activations have already shifted to one or the other side of the activation curve. As a result of this, there is a loss of plasticity associated with learning the early-trained patterns that cannot easily be overcome by differences in cumulative frequency. The latter implies that AoA will have a stronger effect than predicted by the cumulative-frequency hypothesis.

Zevin and Seidenberg (2002) pointed out that this state of affairs will indeed be observed in neural networks, but only when there is little or no overlap between the input–output mappings of the information that is learned first and the mappings that are learned later. Otherwise, the later acquired mappings profit from the early-acquired mappings. Therefore, they predicted a stronger effect of AoA in tasks that involve an arbitrary mapping between input and output (they mentioned the relationship between the orthography and the meaning of words as one possible example), whereas in tasks that allow a carry-over of early learned information to later to be learned information the effect of AoA would reduce to the cumulative-frequency hypothesis (they mentioned the mapping from English orthography to phonology as an example of this type of mapping).

In summary, there are many different explanations of the AoA effect in word processing. The first explanations heavily focused on the idea that the AoA effect had another origin than the frequency effect. The most frequently cited origin of the AoA effect is the phonological output lexicon. Although this view still seems to prevail, it is rapidly being supplemented by views that put less emphasis on different stages in word processing, but instead point to the fact that AoA and frequency effects are both likely to be the result of the way in which information is stored and accessed in the brain. These views predict that AoA will have an effect whenever frequency has an effect (and so that there is no point in looking for task dissociations, such as the distinction between tasks that tap input and output processes), but they differ in the relative weight that has to be attributed to each variable. The cumulative-frequency hypothesis critically rests on the prediction that the weights of frequency and years-known must be the same. The connectionist type of explanation predicts that the weight of AoA will be significantly larger than the weight of frequency, whereas the growing network model at present does not make specific predictions about the relative weights.

4. A multi-task investigation of AoA and frequency effects in visual word processing

It occurred to us that the best way to decide between the different explanations of the AoA effect was to systematically compare the effects of AoA and frequency in a series of tasks that differed on the extent to which they relied on different types of information. If the AoA effect comes from the phonological output lexicon, then it should be stronger for tasks that are believed heavily to rely on information from this system. Indeed, the failures to observe an effect of AoA in perceptual identification (Gilhooly & Logie, 1981) and object categorization (Morrison et al., 1992) have been part of the argument to situate the effect in the later stages of speech output. Similarly, if the AoA effect predominantly comes from the organization of the semantic system, as defended by Brysbaert et al. (2000), then the effect should increase the more the task relies on the meaning of the words. In contrast, if the AoA and frequency effects are both the result of the way in which information is stored and accessed in the brain, then there should be a strong correlation between both effects: The AoA effect should be large in tasks with a big frequency effect, and it should be weak in tasks with a small frequency effect. Finally, the cumulative-frequency hypothesis makes the additional prediction that the weight of the frequency effect and the AoA effect in Eq. (2e) must be the same, as they both represent the exponent of the power function.

Eight visual word processing tasks were selected on the basis of the processes that are believed to underlie successful performance on each task. The first task was *perceptual identification* with tachistoscopic presentation of the stimuli and percentage of recognition as the dependent variable. This task is generally believed to measure the activation of visual word representations, although a response bias is likely to be involved as well (e.g., Broadbent, 1967; Morton, 1979; Ratcliff & McKoon, 2000; Wagenmakers, Zeelenberg, & Raaijmakers, 2000). The task usually shows a reliable frequency effect and is particularly important for checking the locus of the AoA effect, because a previous failure to obtain the effect with this task (Gilhooly & Logie, 1981) has been interpreted as evidence that AoA does not play a role in the organization of the input lexicon.

The second type of task involved word naming. There were three variants of this task: immediate naming, delayed naming, and speeded naming. The *immediate naming* task is the best known and mostly used variant. Participants simply have to read aloud the word that is presented on the computer screen and the voice onset time is registered. However, because the onset time and its registration are heavily influenced by the nature of the first sound of the word (Morrison & Ellis, 2000; Treiman, Mullenix, Bijeljac-Babic, & Richmond-Welty, 1995), researchers usually add a *delayed naming* task to the immediate naming task. In the delayed naming task, participants are given enough time to prepare the response before it has to be emitted, so that any variation in response times must be due to the articulation process and its registration. Finally, a *speeded naming* task was administered, because Gerhand and Barry (1999a) reasoned that forcing the participants to pronounce the word as rapidly as possible would increase the impact of the phonological output part in the response latencies. In line with their expectations, they observed a stronger AoA effect in the speeded naming task (28 ms) than in the immediate naming task (14 ms), while the frequency effect remained the same (26 vs. 22 ms).

The third type of task consisted of three *lexical-decision* tasks. The variants were defined in terms of the non-words, which were either illegal non-words, legal non-words, or pseudohomophones. Stone and Van Orden (1993) examined the word frequency effect for these three types of lexical decisions, and found that the effect was smallest when the non-words were illegal (i.e., largely consisted of strings of consonants) and largest when the non-words were pseudohomophones (i.e., sounded like real words). The task with illegal non-words is of additional interest, because it seems to eliminate the effects of semantic and phonological word characteristics. For example, James (1975) observed a concreteness effect for low-frequency words in a lexical-decision task with legal non-words but not with illegal non-words, suggesting that the decision in the illegal non-word condition did not involve semantic information. Similarly, Schulman, Hornak, and Sanders (1978) and Gibbs and Van Orden (1998) failed to find differences between reaction times to words with consistent letter–sound mappings (e.g., *beech*) and words with inconsistent mappings (e.g., *beard*) in a lexical-decision task with illegal non-words, suggesting that the decision in this task does not rely on phonology (contrary to the conditions with legal non-words and pseudohomophones). The task with pseudohomophones is also of interest, not only because it leads to a larger frequency effect, but because in many models it is believed to suppress the reliance on phonological information (as this information does not make a distinction between the word and the non-word trials; but see Gibbs & Van Orden (1998), for a different interpretation of this task). Gerhand and Barry (1999b) ran these three versions of the lexical-decision task in English and reported frequency effects of 33, 77, and 90 ms in the tasks with illegal, legal and homophonic non-words respectively, together with AoA effects of 22, 59, and 39 ms (and a significant interaction between both variables). Surprisingly (given that they found reliable AoA effects in the conditions with illegal and homophonic non-words), the authors did not consider these data as evidence against their phonological output hypothesis of the AoA effect.

Finally, we included a semantic task in which participants had to decide whether the stimulus that was presented referred to an object (i.e., was a noun) or was the first name of a person. This task has the advantage that all words used in the other experiments can be included in the same category (see Brysbaert et al. (2000) for a discussion of this point). In addition, Taft and van Graan (1998) reported that the task is insensitive to phonology effects, as there was no time difference in the responses to

Table 1
Predictions of the AoA effect in the different experiments, derived from the various theoretical accounts

Experiment	Theoretical account			
	POH	SEM	CFH	Comp. models
Perceptual identification	–	–	AoA = Freq	AoA ≈ Freq
Immediate naming	+	+	AoA = Freq	AoA ≈ Freq
Delayed naming	?	–	AoA = Freq	AoA ≈ Freq
Speeded naming	++	–	AoA = Freq	AoA ≈ Freq
Word/illegal non-words	–	–	AoA = Freq	AoA ≈ Freq
Word/legal non-words	+	+	AoA = Freq	AoA ≈ Freq
Word/pseudohomophones	–	+	AoA = Freq	AoA ≈ Freq
Word/first name	–	++	AoA = Freq	AoA ≈ Freq

POH = phonological output hypothesis; SEM = the semantic hypothesis; CFH = the cumulative-frequency hypothesis; Comp. models = the computational models, either the connectionist models or the growing network model.

words with regular spelling-sound mapping (*plank*) and words with irregular mapping (*pint*). [1]

The predictions with regard to these tasks were rather straightforward and are summarized in Table 1. If AoA only has an effect in the phonological output system, it should be clearly present in the naming tasks and the lexical-decision task with legal non-words, but less so in perceptual identification, lexical decision with illegal and homophonic non-words, and in the semantic noun/first-name categorization task. Alternatively, if AoA has its major role in the organization of the semantic system, its effect should be largest in the semantic noun/first-name categorization task, would probably also be observed in lexical decision with legal non-words and pseudohomophones, could have an effect in immediate naming (assuming semantic involvement in the naming of Dutch words; see Reynvoet, Brysbaert, & Fias, 2002), but would be less expected in perceptual identification, speeded naming, and lexical decision with illegal non-words. In contrast, the three remaining accounts predict a strong correlation between the magnitude of the frequency effects and the magnitude of the AoA effects. Only the cumulative-frequency hypothesis predicts that the weights of both variables should be the same.

The conclusions of our multi-task investigation not only depend on the choice of tasks, but also on the quality of the stimulus materials. For instance, a few years ago one of the present authors (MB) was unable to find an AoA effect in Dutch word naming because the range of AoA values available at the time was too small. Similarly, Zevin and Seidenberg (2002) pointed to the need of very good frequency

[1] We agree that this task may not be the most obvious semantic task. It is, however, virtually impossible to find another semantic task in which we can include the same stimuli as in the other seven experiments, which was the key feature of the present study. Ghyselinck (2002) reported a more "classic" semantic task with a man-made/naturally occurring binary decision on visually presented words. That study resulted in a reliable AoA effect of 49 ms and a frequency effect of 91 ms. For other demonstrations of AoA effects in semantic tasks, see Ghyselinck, Custers, and Brysbaert (in press).

measures for this type of research. In particular, they claimed that the traditional Kučera and Francis (1967) measures for English do not seem to be a good choice, because they are based on 1 million words only, come from a restricted range of texts and are already a few decades old. We made use of the Celex frequency measures (Baayen, Piepenbrock, & van Rijn, 1993) which are based on a corpus of 42,380,000 Dutch words coming from a large variety of sources. As for AoA, we had student estimates for a total of 2816 four- and five-letter words (Ghyselinck et al., 2000), which not only correlated well with judgments of teachers about which words should be known by 6-year old children ($r = 0.80$), but also with the actual knowledge of children at the age of 6 ($r = 0.75$) and at the age of 12 ($r = 0.81$; De Moor et al., 2000). In addition, these AoA estimates had no cut-off values at the lower and the higher end (as is the case for all measures obtained with a rating scale), because we simply asked participants to indicate the age at which they thought they had acquired the words.

5. Method

Stimuli. For all experiments we used the same set of stimuli. The set consisted of 192 words for which we had measures of AoA and frequency. The first half of the set was constructed so that we had an orthogonal variation of word frequency and word AoA. The next set of 96 words was obtained by creating two lists of 24 words each that differed in AoA and were matched on frequency, and two lists of 24 words each that differed in frequency and were matched on AoA. The first set of words was created to examine possible interactions between AoA and frequency by means of analyses of variance. The second to make sure that the effects we found were not due to exception words (due to the strong correlation between frequency and AoA, there are very few words with an early AoA and a low frequency, and a late AoA and a high frequency) and to be able to fully exploit the data by means of regression analyses. For every set the words were matched on word length, number of syllables, and number of orthographic neighbors (see Table 2 for the correlations between these variables). The words could not be fully matched on word imageability and word familiarity without seriously compromising the range of values. Word imageability was not fully controlled in the present experiments because all our previous research in Dutch showed that the effect of this variable in visual word processing is negligible

Table 2
Correlations ($N = 192$) between AoA, Log (Freq), word length, number of syllables, and number of orthographic neighbors

	AoA	Log (Freq)	Word length	Syllables
Log (Freq)	−0.12			
Word length	0.06	−0.07		
Syllables	0.09	−0.01	0.40*	
Neighbors	−0.18*	0.11	−0.40*	−0.33*

*Significant at the $p < 0.05$ level or greater.

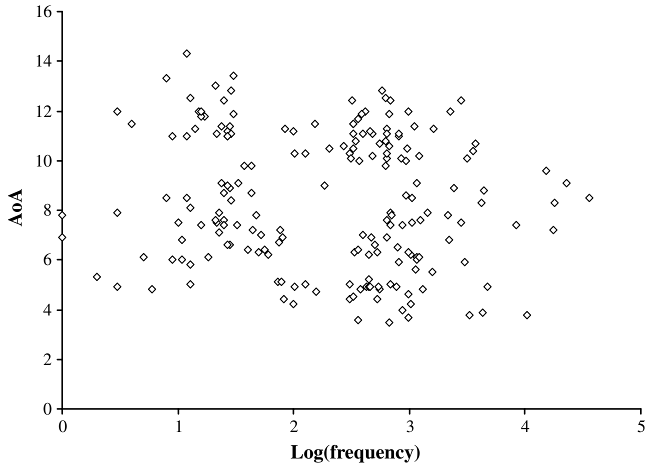

Fig. 1. Scatterplot of the frequency and AoA of the 192 stimuli.

once stimuli are controlled for AoA (see Section 1). Word familiarity was not fully matched because we are convinced that familiarity ratings are largely based on the cumulative frequency with which words have been encountered in the past, so that controlling for this variable actually involves a contradiction. [2] The AoA measures were based on student ratings collected by Ghyselinck et al. (2000). For each word, students estimated the age at which they had learned the word. Frequency measures were based on the Celex Lexical Database (Baayen et al., 1993), which lists frequency counts based on a total of 42,380,000 written words. We worked with the lemma counts and used the base 10 logarithm of the actual frequencies. The selected words as a function of frequency and AoA are shown in Fig. 1. All words, together with their AoA, frequency and first English translation in the Dutch–English dictionary "Van Dale Handwoordenboek Nederlands-Engels" are listed in an Appendix that can be consulted on Elsevier's website. As can be seen in Fig. 1, log frequency ranged from less than 0.5 (below 1 per million) to more than 4.0 (over 230 per million). AoA ranged from slightly below 4 years to over 12 years. In this respect, it may be interesting to note that few participants in the Ghyselinck et al. (2000) rating study thought they had acquired words before the age of 3, which may be an indication of the well-known phenomenon of infantile amnesia (e.g., Howe & Courage, 1993).

Participants. All 161 participants were first-year students in psychology and educational sciences from Ghent University (ranging in age from 17 to 48 years, most of them being between 18 and 20 years). They participated for course credits. All spoke Dutch as their first language, and had normal or corrected to normal vision. For

[2] These a priori considerations were confirmed by post hoc regression analyses. In no task did rated familiarity or imageability explain a significant percentage of variance, in addition to the effects of AoA and frequency.

Table 3
Number of participants, mean reaction latencies (or percentage correct for perceptual identification), mean age (range shown in parentheses) and regression coefficients (by items analyses) for frequency and AoA per experiment (standard errors are shown in parentheses)

Experiment	*N*	Mean(sd)	Age(range)	Coefficients (Std errors)	
				ln(Frequency)	ln(Age – AoA)
Perceptual identification	20	45(19)	23(17–39)	−0.013* (0.006)	−0.301* (0.053)
Immediate naming	21	524(42)	18(17–20)	0.001 (0.007)	−0.204* (0.064)
Delayed naming	17	362(28)	25(18–41)	−0.015 (0.012)	0.041 (0.109)
Speeded naming	23	455(33)	18(17–20)	−0.014[a] (0.008)	−0.163* (0.071)
Word/illegal non-words	20	512(42)	19(17–22)	−0.019* (0.006)	−0.194* (0.049)
Word/legal non-words	20	662(88)	18(17–21)	−0.055* (0.006)	−0.476* (0.049)
Word/pseudohomophones	20	815(156)	20(18–24)	−0.047* (0.007)	−0.468* (0.059)
Word/first name	20	683(92)	19(17–23)	−0.018* (0.006)	−0.245* (0.051)

*Significant at the $p < 0.05$ level or greater.
[a] $p < 0.08$ so can be considered to be significant on a one-tailed test.

each experiment the total number of participants is indicated in Table 3, together with their average age. [3]

Procedure. The eight experiments all followed the same procedure. Participants were given written instructions on the computer screen. Each trial consisted of the presentation of a central fixation point ("+") for 500 ms, followed by a blank interval of 500 ms, and the presentation of the stimulus. All stimuli were presented in a single block, with a different permutation for each participant. Before the experimental block started, each participant finished 20 practice trials. The words (including the first names) were presented in lower-case letters.

Perceptual identification. In this task participants had to identify a word that was presented briefly on the computer screen. The stimulus was shown for 33 ms, followed by a pattern mask. Presentation onset was synchronized with the vertical retrace of the computer screen. The participants told the experimenter what they had seen and the experimenter encoded the correctness of the response on-line via the computer keyboard.

Naming tasks. There were three naming tasks: Immediate naming, delayed naming, and speeded naming. In the first task, participants were asked to name the stimulus immediately upon presentation. The stimulus remained on the screen for 770 ms or until the participant reacted. In the second task, participants were instructed to wait for a go-signal before they started to pronounce the word. In this task, stimulus words were presented for 770 ms followed by the go-signal (two square brackets)

[3] Although the number of participants per experiment might seem rather low, we are confident that the experiments had enough power to reveal the effects that concerned us. In Morrison and Ellis's (1995) seminal paper, the average number of participants per experiment was 18.

after a random time interval between 2000 and 2100 ms. Finally, in the speeded naming task, the stimulus words were shown for 400 ms and participants were encouraged to name the word before it disappeared.

Lexical-decision tasks. We ran three lexical-decision tasks in which participants had to make a distinction between known words and either (1) illegal non-words, (2) legal non-words, or (3) pseudohomophones. Illegal non-words were created by replacing the vowels of existing words by consonants so that the letter string became unpronounceable (e.g., *zrlf*, *wrcnf*). The legal non-words were created by replacing a vowel or a consonant of an existing word by another vowel or consonant that respected the orthographic rules of Dutch (e.g., *zilf*, *treef*). The pseudohomophones also respected the orthographic rules of Dutch but in addition sounded like real words (e.g., *gijt*, *koort*). Participants indicated whether the presented letter string was a word or a non-word by pressing the "q" or "p" key of the keyboard. The stimulus–response mapping was counterbalanced over participants. The letter string remained on the screen until the participant reacted.

Semantic task . In the last task, we asked participants to indicate whether the presented word referred to a noun with a definable meaning or to the first name of a person. The advantage of this task was that all our test words required the same response. Only generally accepted first names were used as foils, with a wide range of frequencies according to the Celex database.

6. Results

Analysis of the orthogonal stimulus set. In a first series of analyses, we ran ANOVAs on the set of 96 words that allowed an orthogonal variation of word frequency and AoA. For all tasks, except for the perceptual identification, errors were removed and analyses were performed without these values. For the perceptual identification task, the dependent variable was the proportion of participants who identified each word correctly. For the other tasks, harmonic means of the RTs were calculated per condition and per participant (or stimulus word). We used this method following Gerhand and Barry's (1999b) and Ratcliff's (1993) suggestions for the most appropriate data transformation in ANOVAs. The results of these analyses for all the experiments are summarized in Table 4.

Perceptual identification. Percentage correct responses was the dependent variable of interest. There was a significant AoA effect of 14.7% [$F1(1,19) = 27.1$, $p < 0.001$; $F2(1,92) = 10.74$, $p < 0.05$] and a significant frequency effect of 9.7% [$F1(1,19) = 22.92$, $p < 0.001$; $F2(1,92) = 4.44$, $p < 0.05$]. There was no interaction effect.

Naming tasks. Response times shorter than 200 ms and longer than 2000 ms were discarded from the analyses. Together with the bad time registrations indicated online by the experimenter, this resulted in a loss of 5.2% of the data in the speeded naming task, 5.0% in the immediate naming task and 5.4% in the delayed naming task. In the speeded naming task both AoA and frequency had significant non-interacting effects. There was a significant 14 ms effect for AoA [$F1(1,22) = 7.71$, $p < 0.05$; $F2(1,92) = 4.4$, $p < 0.05$], and a 9 ms effect for frequency

Table 4
Arithmetic means of the harmonic mean reaction latencies in milliseconds (or percentage correct for perceptual identification) and percentage of errors as a function of AoA and frequency (based on the orthogonal set of 96 stimuli)

Experiment	Early/high	Late/low	Effect
Perceptual identification			
AoA	50	35	15
Frequency	48	38	10
Immediate naming			
AoA	527	544	17
Frequency	531	540	9
Delayed naming			
AoA	370	365	−5
Frequency	361	375	14
Speeded naming			
AoA	454	467	14
Frequency	456	465	9
Word/illegal non-words			
AoA	491 (1.7)	503 (3.9)	12 (1.7)
Frequency	488 (2)	506 (3.6)	18 (1.6)
Word/legal non-word			
AoA	604 (4.2)	679 (22.3)	75 (18.1)
Frequency	607 (5.7)	676 (20.9)	70 (15.2)
Word/pseudohomophones			
AoA	704 (3.3)	820 (12.6)	117 (9.3)
Frequency	711 (3.2)	812.6 (12.7)	102 (9.5)
Word/first name			
AoA	626 (4.1)	678 (6.5)	52 (2.4)
Frequency	635 (1.9)	669 (8.7)	35 (6.8)

[$F1(1,22) = 6.36$, $p < 0.05$; $F2(1,92) = 2.14$, $p > 0.1$]. The immediate naming task only revealed a significant AoA effect [$F1(1,20) = 6.24$, $p < 0.05$; $F2(1,92) = 3.73$, $p > 0.05$], and no significant frequency effect although there was a 9 ms difference between high frequent and low frequent words. A reversed pattern is found for the delayed naming task: There was only a significant 14 ms effect of frequency [$F1(1,16) = 4.65$, $p < 0.05$; $F2(1,92) = 5.12$, $p < 0.05$].

Lexical-decision tasks. In the condition with homophonic non-words one participant was excluded because s/he had made 75% errors to late-acquired, low-frequency words. As can be seen in Tables 3 and 4, the manipulation of the non-words was successful: Reaction times were significantly shorter in the condition with illegal non-words than in the condition with legal non-words, and these in turn were shorter than in the condition with pseudohomophones. Separate 2×2 ANOVAs were run for the three tasks. In all analyses, the same pattern of results emerged: Both AoA and frequency had significant non-interacting effects on decision latencies.

The AoA effect was 12 ms in the condition with illegal non-words [$F1(1,19) = 10.83$, $p < 0.01$; $F2(1,92) = 5.22$, $p < 0.05$], 75 ms in the condition with legal non-words [$F1(1,19) = 74.63$, $p < 0.001$; $F2(1,92) = 56.61$, $p < 0.001$], and 116 ms in the condition with pseudohomophones [$F1(1,18) = 82.72$, $p < 0.001$; $F2(1,92) = 46.72$, $p < 0.001$]. The frequency effects were respectively 18 ms [$F1(1,19) = 14.7$, $p < 0.01$; $F2(1,92) = 11.83$, $p < 0.001$], 70 ms [$F1(1,19) = 69.32$, $p < 0.001$; $F2(1,92) = 45.27$, $p < 0.001$], and 102 ms [$F1(1,18) = 57.12$, $p < 0.001$; $F2(1,92) = 36.23$, $p < 0.001$]. In none of the conditions did the interaction reach significance [illegal non-word condition: $F1(1,19) = 1.26$; $F2(1,92) = 0.66$; legal non-word condition: $F1(1,19) = 0.45$; $F2(1,92) = 1.7$; pseudohomphone condition: $F1(1,18) = 2.23$; $F2(1,92) = 1.67$].

Semantic task. A 2×2 ANOVA was run which revealed that both AoA and frequency had significant non-interacting effects on decision latencies: an AoA effect of 52 ms [$F1(1,19) = 16.79$, $p < 0.001$; $F2(1,92) = 12.41$, $p < 0.001$], and a frequency effect of 35 ms [$F1(1,19) = 19.69$, $p < 0.001$; $F2(1,92) = 11.44$, $p < 0.01$].

Testing the cumulative-frequency hypothesis. To provide a critical test of Lewis's cumulative-frequency hypothesis, the data were transformed according to the method outlined in Lewis (1999a). Frequency scores were log transformed, and AoA scores were replaced by the log of years-known (e.g., $\ln(20 - \text{AoA})$ for a twenty-year-old person). The reaction times (minus 300 ms) were also log transformed, and the number of errors (for perceptual identification) was analyzed in both their raw state and using a transformation that adjusted for skew ($\ln((p(\text{error}) - 1)/p(\text{error}))$). With the exception of those for perceptual identification, all analyses were performed by items and by subjects (see Lorch & Myers, 1990; results are only reported as significant if they are significant by both methods).

The transformed data from all the experiments were subjected to multiple regression analyses (coefficients summarized in Table 3). All four decision tasks show a similar pattern of results with increased log of frequency and increased log of time known predicting decreased reaction times. The perceptual identification task showed the same result for the pattern of errors with an increase in these factors leading to a decrease in number of errors.

The analysis of the naming tasks was not as clear cut as for the other types of tasks although the direction of the results was, for the most part, consistent with those of the lexical-decision tasks. The immediate naming experiment showed a significant effect of log of time known but a non-significant effect of log of frequency. The speeded naming experiment showed a significant effect of log of time known and an almost significant effect of log of frequency ($0.05 < p < 0.10$). Neither predictor was significant for the delayed naming task.

The addition of an interaction term was incorporated into the regression analyses for each of the experiments. Most of these interactions were non-significant and so could be discounted from the analysis (Aiken & West, 1991). The one significant interaction was for the immediate naming experiment ($b = 0.073$; $p < 0.05$) and this changed the significance levels of the log of frequency ($b = -0.176$; $p < 0.05$) and log of time known ($b = -0.596$; $p < 0.05$).

In order to analyze the factors affecting naming retrieval when any effects of name production had been removed, the immediate naming experiment and the delayed naming experiment were analyzed together. The dependent variable was the log of the difference in time required to name each word immediately and to name it after a delay. This was regressed against log of time known and log of frequency. This analysis found a significant effect of time known ($b = -0.387$; $p < 0.05$) and a non-significant effect of frequency ($b = 0.021$; $p > 0.05$). Inclusion of the interaction in this analysis was not, itself, significant and it did not change the significance of the main factors.

The cumulative-frequency hypothesis rests on the prediction that, when time known and frequency are log transformed then they should predict the log of reaction times such that the regression coefficients of each are equal (see Eq. (2e)). Such a comparison can be made in a number of ways and each of these ways provides evidence that the two coefficients are consistently different with those for time known being larger than those for frequency.

Comparison of the coefficients from the by items analyses can be made using the standard errors as estimates (see Table 3). In all experiments (except delayed naming where no significant effects were found) the two coefficients differ by more than twice the maximum standard error. This means that the by items coefficients are significantly (z's > 2; p's < 0.05) different for all seven of the experiments. Similar conclusions can be drawn from the analysis by subjects. Paired (by subjects) t tests were conducted to contrast the size of the two coefficients for each experiment. For all six experiments considered, the time known coefficient was significantly larger than the frequency coefficient (t's$(19) > 3.496$; $p < 0.05$, delayed naming was excluded because no time known effect was found; perceptual identification could not be analyzed by subjects).

Whilst the analyses reported above demonstrate that the contribution of the time known factor is consistently greater than the contribution of the frequency factor, they do not elucidate the nature of the relationship between the two factors across the tasks. In order to investigate how the relative contribution of frequency and time know changes over different tasks it is necessary to perform a global analysis. Such an analysis is shown in Fig. 2, which plots the regression coefficients for time known against the regression coefficients for frequency for each of the eight experiments. This plot shows that there is a roughly linear relationship between the coefficients over the eight experiments ($r(7) = 0.74$; $p < 0.05$). Further, the intercept of regression line through these points is not significantly different from zero ($t(6) = 1.442$; $p > 0.10$) and so we cannot reject the possibility that the regression line passes through the origin. Indeed, a regression conducted with no intercept (i.e., a set to zero) provides an equally good fit to the data ($r(7) = 0.911$; $p < 0.05$). It is wholly possible, therefore, that we would expect to find a frequency effect wherever we find a time known effect, albeit, the frequency effect would be of a magnitude 9.4 times smaller than that for time known effects.

Further analyses indicated that none of the above conclusions depended on the fact that a constant of 300 ms had been subtracted from the RTs (see the parameter a in Eq. (2)). The same results were obtained with other estimates of the parameter

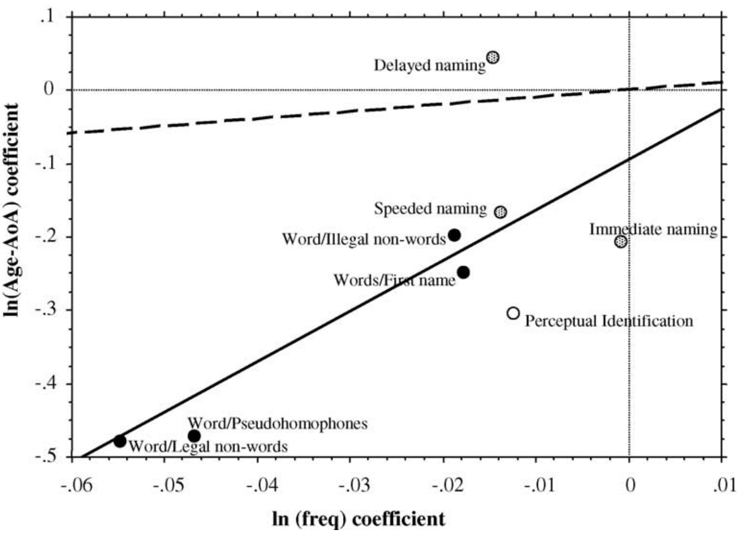

Fig. 2. By items coefficients for the eight experiments represented in a two dimension scatterplot. Lexical-decision or semantic tasks are shown as black points, naming tasks are shown as grey points and the identification task is shown as an open point. The bold line through the plot shows the correlation between the coefficients and the dashed line represents where this correlation line should be for the data to be consistent with the cumulative-frequency hypothesis.

(see also Lewis et al., 2001). We used the estimate of 300 ms because it was well below the fastest RTs in our sample (the RTs in the present experiments tended to be faster than those in comparable English experiments, in particular for the naming tasks).

7. Discussion

In the last five years, it has become clear that the age at which a word is learned is an important determinant of the ease with which a word is recognized in a variety of visual word processing tasks. While doing research on this topic, many authors have been guided by the assumption that the origin of the AoA effect was very narrow and could be pinned down to one particular stage of the word processing chain. Some of them even maintained this assumption, despite the finding that AoA effects popped up in several different tasks where they were not expected. For instance, Gerhand and Barry (1999a, 1999b) interpreted small differences in the impact of AoA and frequency in five different tasks (lexical decision with illegal, legal, and homophonic non-words, plus immediate and speeded naming) as evidence for the hypothesis that the AoA effect originated from the phonological output lexicon, even though the most conspicuous results of their studies were (1) the presence of a reliable AoA effect in all different tasks, even in those that on the basis of previous research were known not to require phonological information to make the binary decision, and (2) a correlation of 0.80 between the size of the frequency effect and the size of the

AoA effect in the different tasks (see Section 1 for the values of the five data pairs). Similarly, many authors have put a lot of weight on two rather old studies that failed to obtain an AoA effect (Gilhooly & Logie, 1981; Morrison et al., 1992).

To further sort out the status of AoA in visual word recognition, we ran eight tasks with the same set of stimuli. In this way, each study had a comparable power to reveal the effects. In addition, we made sure that we had a large range of frequency and AoA values, and we optimized the analysis by looking at the regression weights calculated on the basis of individual reaction times at the item level. Our main aim was to find out whether we would find a pattern of regression weights that was in line with a narrow origin of the AoA effect, or whether we would be confronted with an AoA effect that was as broad as the frequency effect. In case we obtained the latter finding, an additional question was whether the weights of the AoA and the frequency effects would be more in line with the cumulative-frequency hypothesis or with the plasticity hypothesis based on connectionist modeling.

The results turned out to be rather straightforward. First, just like Gerhand and Barry (1999a, 1999b), we obtained a strong linear relationship between the magnitude of the frequency and the magnitude of the AoA effect in the different tasks, making it difficult to maintain that they do not have a common basis. Looking at Fig. 2, it is clear that focussing on the deviations from the linear regression line while ignoring the very existence of the regression itself largely would be missing the point (as indeed we argue Gerhand and Barry have done). Second, the weight of the AoA parameter (i.e., number of years-known) is so much bigger than the weight of the frequency factor that it becomes very difficult to keep on defending the cumulative-frequency hypothesis, even though the cumulative-frequency hypothesis was right in its prediction of a lack of interaction between AoA and frequency. Take, for instance, the word/illegal-non-word decision task. RTs in this task are predicted by the following equation:

$$\mathrm{Ln}\,(\mathrm{RT_{LD}} - 300) = 5.747 - 0.019\ln(\mathrm{freq}) - 0.194\ln(\text{years-known}) \tag{3a}$$

$$\mathrm{RT_{LD}} - 300 = \mathrm{e}^{(5.747 - 0.019\ln(\mathrm{freq}) - 0.194\ln(\text{years-known}))} \tag{3b}$$

$$\mathrm{RT_{LD}} = 300 + \mathrm{e}^{5.747} * (\mathrm{freq})^{-0.019} * (\text{years-known})^{-0.194} \tag{3c}$$

$$\mathrm{RT_{LD}} = 300 + 313 * [(\mathrm{freq}) * (\text{years-known})^{10}]^{-0.194} \tag{3d}$$

A very similar equation could be written for each of the tasks. In particular, there will always be an extra exponent added to the variable years-known, which will not be present for the frequency variable. The value of this exponent will always be around 10, meaning that the impact of the variable years-known is about 10 times the impact of the variable word frequency. It is important to realize that this extra exponent is not due to the time unit or to the frequency unit chosen. It remains there irrespective of whether years-known is expressed in years or in months, and irrespective of whether the frequency is measured per 42 million, per million, or per thousand (the only value that will change in these examples is the value of 5.747 in Eq. (3a)).

To understand the importance of the extra exponent in Eq. (3d) for the cumulative-frequency hypothesis, it may be good to look at the data in Table 4 (which represent the effects for the stimulus set with the orthogonal manipulation of AoA and frequency). Here we see, for instance, that the effects of AoA (years-known) and frequency are nearly the same in the word/legal-non-word lexical-decision task: 75 vs. 70 ms. Yet, the earliest learned items could not have a value of years-known much above 18 (i.e., a participant of 20 years and a word acquired at the age of two) and the latest learned items could not have a value of years-known much less than 2 (i.e., a participant of 20 years and a word acquired at the age of 18). So, the "oldest" words were at most 9 nine times "older" than the youngest and thus 9 times more often encountered. The frequencies, on the other hand, ranged from some 2 per million to some 180 per million (see Fig. 1), so that the high-frequency words were encountered on average 90 times more often than the low-frequency word. Despite this huge difference in range, the magnitudes of the AoA and the frequency effect were the same. This cannot be explained by the cumulative-frequency hypothesis, independent of which mathematical equation is used; hence, the extra exponent of 10 for the AoA variable.

Given the magnitude of the extra exponent, we fear that it will not be easy to find ways to salvage the cumulative-frequency hypothesis. For instance, Zevin and Seidenberg (2002) hypothesized that the Celex frequency estimates are biased towards adult texts and therefore underestimate the cumulative frequency of words that figure predominantly in childhood years. Although this may indeed explain some of the AoA effect, the extra exponent of Eq. (3d) implies that the frequency of the early-acquired words must have been underestimated by a factor of 10. In addition, to get the full picture one probably would have to take into account that the rate of information take-up is likely to be slower in childhood than in adolescence and early adulthood (at least one would hope so for university students), so that one risks to overestimate the cumulative frequency of early words if early frequency estimates are given the same weight as estimates of later frequencies.

Another factor that has been invoked to explain part of the AoA effect, is word familiarity (Zevin & Seidenberg, 2002). Ever since the work of Gernsbacher (1984), it is known that certainly for low-frequency words, the objective frequency counts do not always seem to match the subjective feelings of familiarity, as some of the low-frequency words elicit a reliably higher feeling of familiarity than others. However, in our view, time is ripe to scrutinize the familiarity ratings as a function of what we know about age of acquisition. Maybe some low-frequency words look reasonably familiar, because they were acquired early? Evidence for this hypothesis is found when we take the 136 words of Morrison et al. (1997) for which we have objective measures of AoA (see Section 1; 75% criterion), objective measures of frequency (Baayen et al., 1993), and familiarity ratings (Balota, Pilotti, & Cortese, 2001). For this set of stimuli, with objective and independent measures, we find a correlation of 0.54 between the familiarity ratings and log frequency, and a correlation of −0.41 between familiarity and the objective AoA measures. In addition, frequency and AoA together accounted for 16% of the variance in lexical-decision times of students (Balota et al., 1999), against 14% of the variance explained by familiarity. So, rather than taking the ill-defined familiarity ratings as the basis of our word-modeling efforts, it seems more sensible to incorporate AoA in our models.

Even though one can never exclude the possibility that in the future a completely different explanation will be found for our findings, at present we think the most sensible approach is to keep in mind that our results were actually predicted by one type of model that has been proposed to account for the AoA effect, and can easily be incorporated in another model. The connectionist models based on a loss of plasticity (Ellis & Lambon Ralph, 2000; Smith et al., 2001) predicted a significantly larger weight for the AoA variable than predicted on the basis of the cumulative-frequency hypothesis, and such a difference would also seem to be in line with Steyvers and Tenenbaum's (submitted for publication) growing network model. So, before dismissing the present findings as another artefact, it may be more worthwhile first to find out what the parameters in these models must look like to yield simulation data similar to our human data. On the basis of these simulations we may get a better understanding of the interplay between time of introduction and frequency of occurrence in the training of learning organisms.

Acknowledgements

This research was made possible by a BOF-project granted to the authors by the Research Council of Ghent University. We thank Johan Pauwels for his help in programming and running the experiments.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. London: Sage.

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

Balota, D. A. (1994). Visual word recognition: The journey from features to meaning. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 303–358). San Diego, CA: Academic Press.

Balota, D. A., Cortese, M. J., & Pilotti, M. (1999). *Item-level analyses of lexical decision performance: Results from a mega-study. Abstracts of the 40th annual meeting of the Psychonomics Society*. Los Angeles, CA: Psychonomic Society, p. 44.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory and Cognition, 29*, 639–647.

Baluch, B., & Besner, D. (2001). Basic processes in reading: Semantics affects speeded naming of high-frequency words in an alphabetic script. *Canadian Journal of Experimental Psychology, 55*, 63–69.

Barry, C., Hirsch, K. W., Johnston, R. A., & Williams, C. L. (2001). Age of acquisition, word frequency, and the locus of repetition priming of picture naming. *Journal of Memory and Language, 44*, 350–375.

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 50*, 560–585.

Bonin, P. (2001). Age of acquisition and word frequency in spoken and written production. Paper presented at the *Conference of the European Society for Cognitive Psychology and the annual conference of the British Psychological Society Cognitive Psychology Section, Edinburgh.*

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology, 93*, 89–114.

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review, 74*, 1–15.

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 and Cognition, 15*, 208–216.

Brysbaert, M. (1996). Word frequency affects naming latency in Dutch with age of acquisition controlled. *European Journal of Cognitive Psychology, 8*, 185–193.

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 of words is a significant variable in semantic tasks. *Acta Psychologica, 104*, 215–226.

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.

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.

De Moor, W., Ghyselinck, M., & Brysbaert, M. (2000). A validation study of the age-of-acquisition norms collected by Ghyselinck, De Moor and Brysbaert. *Psychologica Belgica, 40*, 99–114.

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.

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

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 and frequency effects in speeded word naming. *Cognition, 73*, 27–36.

Gerhand, S., & Barry, C. (1999b). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. *Memory & Cognition, 27*, 592–602.

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.

Ghyselinck, M. (2002). *Age of acquisition as an organizing variable of the semantic system*. Unpublished doctoral dissertation, Ghent University.

Ghyselinck, M., Custers, R., & Brysbaert, M. (in press). The effect of age of acquisition in visual word processing: Further evidence for the semantic hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Ghyselinck, M., De Moor, W., & Brysbaert, M. (2000). Age-of-acquisition ratings for 2816 Dutch four- and five-letter nouns. *Psychologica Belgica, 40*, 77–98.

Gibbs, P., & Van Orden, G. C. (1998). Pathway selection's utility for control of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1162–1187.

Gilhooly, K. J., & Gilhooly, M. L. M. (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 1944 words. *Behavior Research Methods & Instrumentation, 12*, 395–427.

Gilhooly, K. J., & Logie, R. H. (1981). Word age-of-acquisition and visual recognition thresholds. *Current Psychological Research, 1*, 215–225.

Gilhooly, K. J., & Watson, F. L. (1981). Word age-of-acquisition effects: A review. *Current Psychological Research, 1*, 269–286.

Howe, M., & Courage, M. L. (1993). On resolving the enigma of infantile amnesia. *Psychological Bulletin, 113*, 305–326.

James, C. T. (1975). Role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 104*, 130–136.

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

Lewis, M. B. (1999a). Age of acquisition in face categorisation: Is there an instance-based account? *Cognition, 71*, 23–39.

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

Lewis, M. B., Gerhand, S., & Ellis, H. D. (2001). Re-evaluating age-of-acquisition effects: Are they simply cumulative-frequency effects? *Cognition, 78*, 189–205.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.

Lorch, R. F., & Myers, J. L. (1990). Regression analysis of repeated measures data in cognitive psychology. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 146–157.

Milgram, S. (1967). The small-world problem. *Psychology Today, 1*, 60–67.

Moore, V., & Valentine, T. (1998). Naming faces: The effect of AoA on speed and accuracy of naming famous faces. *The Quarterly Journal of Psychology, 51*, 458–513.

Moore, V., & Valentine, T. (1999). The effects of age of acquisition on processing famous faces and names: Exploring the locus and proposing a mechanism. In M. Hann & S. Stoness (Eds.), *Proceedings of the 21st annual meeting of the Cognitive Science Society* (pp. 416–421). Vancouver: Lawrence Erlbaum Associates.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (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, 50*, 528–559.

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*, 116–133.

Morrison, C. M., & 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., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition, 20*, 705–714.

Morrison, C. M., Hirsh, K. W., Chappell, T., & Ellis, A. W. (2002). Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *European Journal of Cognitive Psychology, 14*, 435–459.

Morton, J. (1979). Word recognition. In J. Morton & J. C. Marshall (Eds.), *Psycholinguistics series: Vol. 2. Structures and processes* (pp. 107–156). London: Paul Elek.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). *The University of South Florida free word association norms*. Retrieved October 28, 2003, from University of South Florida Web site: http://w3.usf.edu/FreeAssociation.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the power law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale: Erlbaum.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510–532.

Ratcliff, R., & McKoon, G. (2000). Modeling the effects of repetition and word frequency in perceptual identification. *Psychonomic Bulletin and Review, 7*, 713–717.

Reynvoet, B., Brysbaert, M., & Fias, W. (2002). Semantic priming in number naming. *Quarterly Journal of Experimental Psychology, 55*, 1127–1139.

Schaie, K. W. (1990). The optimization of cognitive functioning in old age: Predictions based on cohort-sequential and longitudinal data. In P. B. Baltes & M. M. Baltes (Eds.), *Successful aging: Perspectives from the behavioral sciences* (pp. 94–117). New York: Cambridge University Press.

Schulman, H. G., Hornak, R., & Sanders, E. (1978). The effects of graphemic, phonetic, and semantic relationships on access to lexical structures. *Memory and Cognition, 6*, 115–123.

Segui, J., & Grainger, J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime-target frequency. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 65–76.

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* (pp. 52–58). Cambridge, MA: MIT Press.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8*, 411–416.

Sternberg, S. (1969). The discovery of processing stages: Extension of Donders' method. *Acta Psychologica, 30*, 276–315.

Steyvers, M., & Tenenbaum, J. (submitted for publication). *The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth*. Retrieved October 28, 2003, from University of California, Department of Cognitive Sciences Web site: http://www-psych.stanford.edu/~msteyver/papers/smallworlds.pdf.

Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 744–774.

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.

Taft, M., & van Graan, F. (1998). Lack of phonological mediation in a semantic categorization task. *Journal of Memory and Language, 38*, 203–224.

Treiman, R., Mullenix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology General, 124*, 107–136.

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 and Cognition, 26*, 1282–1291.

van Loon-Vervoorn, W. A. (1989). *Eigenschappen van basiswoorden*. Lisse: Swets and Zeitlinger.

Wagenmakers, E. J. M., Zeelenberg, R., & Raaijmakers, J. G. W. (2000). Testing the counter model for perceptual identification: Effects of repetition priming and word frequency. *Psychonomic Bulletin and Review, 7*, 662–667.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440–442.

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*, 407–421.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language, 47*, 1–29.