COGNITION

ELSEVIER

Cognition 78 (2001) 189–205

www.elsevier.com/locate/cognit

Re-evaluating age-of-acquisition effects: are they simply cumulative-frequency effects?

Michael B. Lewis*, Simon Gerhand, Hadyn D. Ellis

School of Psychology, Cardiff University, P.O. Box 901, Cardiff CF1 3YG, UK

Received 4 July 2000; accepted 8 September 2000

Abstract

The time it takes to read or produce a word is influenced by the word's age of acquisition (AoA) and its frequency (e.g. *Quarterly Journal of Experimental Psychology* 12 (1973) 85). Lewis (*Cognition* 71 (1999) B23) suggested that a parsimonious explanation would be that it is the total number of times a word has been encountered that predicts reaction times. Such a cumulative-frequency hypothesis, however, has always been rejected because the statistical effects of AoA and frequency are additive. Here, it is demonstrated mathematically that the cumulative-frequency hypothesis actually predicts such results when applied to curvilinear learning. Further, the data from four influential studies (two of which claim support for independent effects of AoA and frequency) are re-analyzed to reveal that, in fact, they are consistent with a cumulative-frequency hypothesis. The conclusion drawn is that there is no evidence with which to refute the most parsimonious of explanations, i.e. cumulative frequency can account for both frequency and AoA effects.

Keywords: Re-evaluating; Age-of-acquisition effects; Cumulative-frequency effects

1. Introduction

There exists a large literature reporting investigations aimed at understanding how we perform lexical and phonological tasks involved in speaking and reading. Various factors have been investigated as possibly affecting the speed of these tasks. Some of these are: the word's length; its similarity to other words; its imageability;

* Corresponding author. Fax: +44-1222-874858.
E-mail address: lewismb@cardiff.ac.uk (M.B. Lewis).

PII: S0010-0277(00)00117-7

its concreteness; and its phonological complexity. The research reported here examines just two such possibilities: the frequency with which a word occurs in every day language; and the age at which a word is typically first learnt (known as age-of-acquisition or AoA).

We begin by briefly summarizing the research into the two effects of AoA and frequency and then describe a simple and parsimonious account of these two factors based on cumulative frequency. This account, however, has always been rejected on the basis of additivity of the effects. The alternative accounts of AoA effects are discussed and evaluated. The aim of this research is to demonstrate how the cumulative-frequency hypothesis, in fact, is consistent with all of the empirical evidence. The introduction, therefore, is followed by a description of how the cumulative-frequency hypothesis can actually predict additive effects of AoA and frequency by applying a simple and intuitive revision. It is then considered whether predictions from the revised cumulative-frequency hypothesis are consistent with the empirical data from a range of studies.

1.1. Frequency and/or age of acquisition effects

The first suggestion that the age at which a word is learned (or AoA) affects performance came from a study of aphasic naming by Rochford and Williams (1962). In an attempt to investigate the proposal that aphasic speech represents a regression to an earlier stage of language development, they found that the age at which 80% of children could successfully name an object predicted the proportion of aphasics who could also name that object. The possible significance that word frequency has on reading latencies was first reported in the seminal paper by Oldfield and Wingfield (1965). They demonstrated that words with high frequency were recognized more quickly than those with low frequency. It was subsequently questioned, however, whether the effects they had found were really AoA effects rather than frequency effects. It was suggested that high-frequency words are also acquired early and it is the word's early acquisition that leads to the effects observed by Oldfield and Wingfield. Carroll and White (1973) investigated whether AoA affects the speed with which adults can name pictures of objects. They found that AoA, and *not* frequency, was the main predictor of picture-naming times such that words that are learnt earlier in life are responded to more quickly than later-acquired words. This experiment first raised the whole issue of whether AoA or frequency is the key underlying variable in lexical performance: a question that has been hotly debated ever since.

Following the Carroll and White (1973) experiment, a large number of investigations have used a wide range of tasks in order to investigate the existence and relative importance of the factors of AoA and frequency (e.g. Barry, Morrison, & Ellis, 1997; Brown & Watson, 1987; Brysbeart, 1996; Coltheart, Laxon, & Keating, 1988; Ellis & Morrison, 1998; Feyereisen, van der Borght, & Seron, 1988; Gerhand & Barry, 1998; Gilhooly & Logie, 1982; Hirsh & Ellis, 1994; Lachman, Shaffer, & Hennrikus, 1974; Nickels & Howard, 1995; Snodgrass & Yuditsky, 1996; Yamazaki, Ellis, Morrison, & Lambon-Ralph, 1997). All these studies either found that the AoA of a word or the frequency of a word affects subjects' performance on the word-specific tasks such as picture-naming or reading. Some of the studies, indeed, found that both AoA and frequency affected performance. Frequency effects and AoA effects are obviously important determinants of human performance in word-related tasks and so it is important that the mechanisms by which these factors influence performance are explained.

1.2. The cumulative-frequency hypothesis

Finding effects for both the age at which a word is typically acquired and the frequency with which it occurs has initial intuitive appeal. This appeal comes from the fact that each of these factors affects the total number of times a word is encountered (seen or heard). The higher the frequency of a word, the more it will be encountered; likewise, words that have been known for longer will have also been perceived more often. It would be parsimonious if the two different factors of AoA and frequency could each be accounted for in terms of the total number of times a word has been encountered. Such an account will be referred to as the cumulative-frequency hypothesis.

The cumulative-frequency hypothesis has been considered at various times during the research on AoA and frequency effects. Indeed, Carroll and White (1973) considered it as an explanation for their findings. This parsimonious hypothesis, however, has been repeatedly refuted by a range of evidence, some of which is quite specious. Alternative accounts for AoA effects, therefore, have been offered, some of which are described below.

1.3. Accounting for independent effects of frequency and AoA

The initial reason why the cumulative-frequency hypothesis was rejected as an account of AoA effects was the failure to find a statistical interaction between AoA and frequency effects. It was theorized (by Carroll & White, 1973, and by others) that if frequency effects and AoA effects were each caused by cumulative frequency then an interaction effect should also be found. The reasoning is that words that are known for longer will have had more time over which frequency will have had an effect. With the exception of a recent study by Gerhand and Barry (in press), no interactions between AoA and frequency have been reported. A result of these failures to find a robust interaction has been that AoA and frequency have been assumed to be independent effects and so must be accounted for separately.

Accounting for frequency effects is relatively straightforward and, indeed, there exists qualitative predictions for these phenomena. The predictions come from the Newell and Rosenbloom (1981) *power law of practice*, which predicts that the time taken to do something is a power function of the number of times it has been performed before. Therefore, reading a high-frequency word or naming a high-frequency object will be faster than for their low-frequency counterparts simply because the participant will be further along the power curve of practice. This relationship, of course, is merely descriptive but mechanisms have been suggested that can account for the observed effects. One possible mechanism for such a pattern of results was suggested by Gilhooly and Gilhooly (1979). They invoked the Morton (1969) concept of logogens to account for the frequency effects, where each time a logogen is activated, its threshold is reduced and so logogens of high-frequency words come to have lower thresholds than those of low-frequency words. The lower thresholds would mean that the higher-frequency words are recognized faster.

Explanations for the effects of AoA have been varied and have come from different sources. Carroll and White (1973) originally speculated that memories are stored in a push-down manner where older memories are deeper. Their account, however, was more descriptive than predictive and offers no mechanism for the nature of verbal memories. An alternative explanation based on cerebral lateralization was tested by Ellis and Young (1977) (derived from the ideas by Gazzaniga, 1970). Their experimentation, however, failed to support this lateralization hypothesis.

In their review of the area, Gilhooly and Watson (1981) concluded that the weight of evidence favoured an interpretation that placed frequency at the input logogens and AoA at the output logogens. The theory was that activation of the input logogen would be more likely following repeated activation and, hence, lower thresholds, whereas the time it takes to activate the output logogen would be affected by how easy a word was to assemble. This idea was further developed by Brown and Watson (1987), who advanced the phonological-completeness hypothesis as an explanation for AoA effects. In this account, the mechanism underlying AoA is proposed to be the nature of the stored phonological representations of words in the speech production system. The phonological forms of later-acquired words need to be assembled for pronunciation, whereas those of early-acquired words have complete representations, and may be retrieved directly from the speech output lexicon (although, we are not told how this assembly would take place). As early-acquired words have the advantage of more rapid phonological assembly than later-acquired words, they are easier to produce as a picture-naming or oral-word-reading response. This account, however, does have the problem that it places the locus of AoA effects at the level of phonological retrieval. It has been demonstrated that AoA has a clear effect on the lexical-decision task (i.e. a task that does not *necessarily* require the retrieval of phonology). This account is also inconsistent with the finding of AoA effects in non-phonological face-related tasks (Moore and Valentine (1998) and Lewis (1999) report finding AoA or AoA-like effects in face-recognition and face-categorization tasks, respectively).

1.4. The cumulative-frequency hypothesis – revised

Here, we would like to champion the maligned cumulative-frequency hypothesis as an explanation for AoA and frequency effects. We will describe how the interpretation of the evidence accumulated against it may be flawed and how the same evidence is, in fact, predicted by a simple revision to the hypothesis. We suggest, therefore, that rejection of the cumulative-frequency hypothesis has been somewhat premature.

The revision to the cumulative-frequency hypothesis considered here was originally offered by Lewis (1999) and involves the application of the power law of practice (Newell & Rosenbloom, 1981).[1] This revision, as will be explained, can demonstrate that the interactive effects of frequency and AoA can consistently evoke meretricious additive results.

Lewis (1999) proposes an instance-based model that suggests that the number of instances encountered affects reaction times as a negative power function. This curvilinearity means that the apparent additivity seen between AoA and frequency can be seen as being consistent with the cumulative-frequency hypothesis. The account is based on two assumptions. First, that learning takes place by the accumulation of instances. The total number of instances (n_i) of a particular item will be determined by its frequency (freq_i) and how long it has been known. The length of time something has been known can be expressed as the difference between when the item was first encountered (AoA_i) and the age at testing (Age). This first assumption can be expressed as

$$n_i = \mathrm{freq}_i(\mathrm{Age} - \mathrm{AoA}_i) \tag{1}$$

This simple model assumes that no decay takes place – so knowing something for longer always leads to more instances and better retrieval. The question of how the account handles decay of memories has been dealt with by Lewis (1999) and will not be repeated here. As words do not tend to fall out of usage over short periods the question of decay can be ignored for the present discussion.

The second assumption is that the reaction time to a stimulus (RT_i) is a power function of the number of instances. This is effectively the power law of practice and can be expressed as

$$\mathrm{RT}_i = k(n_i)^{-A} \tag{2}$$

where k and A are free parameters. It is possible to substitute Eq. (1) into Eq. (2) as the term n_i. The subsequent equation is the prediction of a cumulative-frequency hypothesis. In order to investigate this hypothesis, it is useful to use linear relationships between variables. It is possible to convert the hypothesis into a linear form by taking the logarithm of the equation. The resulting hypothesis is of the form

$$\ln(\mathrm{RT}_i) = -A\ln(\mathrm{freq}_i) - A\ln(\mathrm{Age} - \mathrm{AoA}_i) + \ln(k) \tag{3}$$

Using this non-linear hypothesis, the appropriate dependent variable for the analysis is the log-transformed reaction times. It is a property of logs, however, that the log of a product of two factors is the same as the sum of the logs of those factors. It would be the case, therefore, that additivity on the log-transformed factors would imply multiplicity.

As analyses such as ANOVA are robust over many transformations (see Ratcliff, 1993), Eq. (3) would also predict additivity for AoA and frequency effects on raw reaction times. This means that the revised cumulative-frequency hypothesis

[1] Recently, the power law of practice has been questioned regarding its application to individual data by Heathcote, Brown, and Mewhort (2000). They show that, although averaged reaction time data tend to follow a power law, individual learners display patterns closer to an exponential law of practice. The arguments offered here are valid for either a power law or an exponential law and so, as we are considering averaged data we employ the established power law of practice.

predicts additive effects of frequency and AoA but it would also predict that a log-transformed analysis would be more appropriate than an analysis on raw data.

The use of multiple linear regression on data that fit Eq. (3) can also lead to the conclusion of additive factors. This can be most easily demonstrated using hypothetical data. Assume we have two sets of words each of which take frequency values from 1 to 1000. One of these sets has been known for 10 years, whereas the other has only been known for 1 year. By applying the power law of practice with a constant of 0.01 the two sets of data have the appearance shown in Fig. 1. Fitting a linear regression to each of the two data sets produces two near parallel lines. A multiple regression conducted on these data gives significant effects of AoA ($t(1996) = 50.490$, $P < 0.05$) and frequency ($t(1996) = 53.110$, $P < 0.05$) but no interaction ($t(1996) = 0.905$, $P > 0.05$). This could lead to the conclusion that the effects of AoA and frequency are additive but we know that for these hypothetical data the relationship is multiplicative as defined in Eq. (1).

The formulation in Eq. (3) demonstrates that the revised cumulative-frequency hypothesis actually predicts an additive relationship between a function of frequency and a function of AoA as has often been found. Simple additivity of AoA and frequency is not sufficient evidence to reject the revised cumulative-frequency hypothesis as it has often been taken to be. The cumulative-frequency hypothesis can actually predict additive results that do not mean independent effects. It is further predicted that analyses based on log transformations will be more appropriate than those based on raw data. This is investigated below by re-analyzing published data.

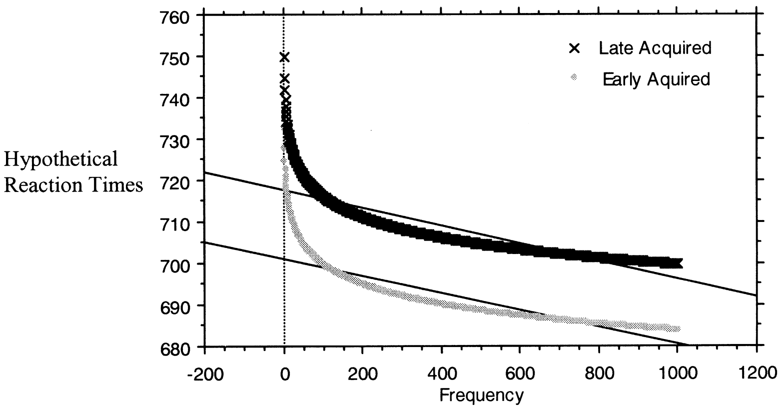

Fig. 1. Hypothetical data based on the revised cumulative-frequency hypothesis. Items vary in their frequency from 1 to 1000 and take values of being learnt 1 year ago (late-acquired) or 10 years ago (early-acquired). The reaction time for each item is calculated using Eq. (3) with $A = 0.01$. The two simple regression lines show the linear correlation between reaction time and frequency for early- and late-acquired items. These lines are nearly parallel indicating that the two effects of frequency and AoA are additive.

2. Re-analysis of published AoA and frequency data

In order to re-evaluate the research conducted on AoA and frequency effects it was necessary to reconsider the data obtained from these studies. It would be, of course, useful to use the raw data from the studies for the re-analysis but this is generally unavailable. What is available, in some cases, are the mean reaction times for the specific items. This means that the analyses offered here are only by-item analyses and not by-subject analyses (see Clarke, 1973). This is not a serious problem for the current study as we are attempting to demonstrate how the original analyses are problematic. In order to do this, it is sufficient to demonstrate that they are problematic either by subjects or by items. Below it will be shown how the original by-items analyses of two important studies are problematic and alternative analyses will be offered.

Of the range of studies that investigate the issues concerned here, only a handful provide by-item reaction times from which it is possible to conduct a new analysis. The limited number of these studies means that it is feasible to consider each one individually. Two key studies of the effects of AoA and frequency were re-analyzed. These studies included one of the first investigations into AoA effects and one of the most recent investigations. These two studies consisted of one factorial design experiment and one regression design using different tasks. A third and fourth study were also re-analyzed. These were a lexical-decision study and a word-naming study that did not originally explore the effects of AoA. By comparing these data sets with AoA and frequency norms it was possible to conduct a much larger test of the cumulative-frequency hypothesis.

2.1. Factorial design: the Gerhand and Barry (1998) data

Within a factorial design, the variables of interest are controlled independently. For experiments involving words or pictures, this procedure requires identifying sufficient exemplars in all possible cells of the factorial design. This is particularly difficult when effects such as AoA and frequency are considered independently because these tend to be correlated (late-acquired words tend to have a low frequency). It is this correlation, it is argued, that gives factorial designs their power over regression designs. By controlling for the two effects independently within a set of stimuli, it is possible to determine the existence of each effect independently from the other.

A recent study by Gerhand and Barry (1998) investigated the effects of AoA and frequency and naming latencies to a range of words. They reported four experiments altogether but by-item data were presented for only the first of these, the word-naming experiment. The last three experiments were not re-analyzed because they produced no AoA or frequency effects, they did not use a factorial design, or they did not explore simple reaction times, respectively. In their first experiment, participants were required to read aloud a presented word. The dependent variable was the time until the participant started to vocalize the word. The original analysis was a factorial ANOVA on the harmonic means of the participants' reaction times and this analysis found a significant effect of frequency ($F(1,60) = 9.61$, $P < 0.01$) and an AoA effect that approached significance ($F(1,60) = 3.235$, $P < 0.1$) but found the interaction to be non-significant ($F(1,60) < 1$).

Casual inspection of the cell means from the Gerhand and Barry (1998) data would lead one to conclude that the effects of AoA and frequency are additive and, hence, by invoking the additive-factors method, operate at different levels (i.e. word production and word recognition, respectively). As demonstrated above, however, this is not sufficient evidence to conclude independent effects. Further, this conclusion is based on an analysis in which the underlying assumptions have been violated. The consequences of this and a method of dealing with these violations are now described.

2.1.1. Alternative log-linear analysis

Closer inspection of the by-item data for Experiment 1 in Gerhand and Barry (1998) reveals a clear relationship between the four cell means and the four cell standard deviations. The correlation between these properties is 0.925, which can be taken to imply that the data have non-homogeneous variances (see Table 1). This concern is further reinforced by the fact that the distribution of the data fails to fit the normal distribution assumed in this type of analysis (data are highly skewed, skew = 1.223). Further, a Cochran's C-test for heterogeneity of variance – a test that determines whether there is a difference between one large variance and the other smaller variances (Kanji, 1993) – demonstrates that the data are indeed significantly heterogeneous ($C(15,4) = 0.480$, $P < 0.05$).

It is often the case that problems of heterogeneity of variance are ignored, particularly following the work conducted by Ratcliff (1993) showing how ANOVA is robust to such violations of assumption. We believe, however, that the violation of the assumption occurs here because the model being used to describe the data is inappropriate. Gerhand and Barry (1998) used a linear relationship between reaction times and AoA and log of frequency. It is possible that the violation of the assumption of heterogeneity of variance is due to this inappropriate model. In order to explore this, we investigate the same data using the model suggested by the revised cumulative-frequency hypothesis.

Table 1
Mean, SD and tests of heterogeneity for the Gerhand and Barry (1998) by-items data and log-transformed data

AoA	Frequency	By-items reaction times		Log (reaction times − 400)	
		Mean	SD	Mean	SD
Early	High	493	24.6	4.501	0.265
Early	Low	518	31.3	4.746	0.260
Late	High	508	26.6	4.653	0.257
Late	Low	533	46.0	4.847	0.317
r (mean, SD)		0.930		0.664	
Cochran's C(15,4)		0.480, $P < 0.05$		0.329, $P = 0.62$	

The revised cumulative-frequency hypothesis suggests that analyses should be conducted on logarithms of reaction times in order to investigate non-linear effects. It can be seen that such a log transformation of reaction times also reduces the heterogeneity of variance and produces a more normal distribution of the Gerhand and Barry (1998) reaction time data. A log transformation of the reaction times minus 400 ms was found to be most appropriate[2] (a simple log transformation also gave the same results). The ANOVA was conducted on the transformed reaction times with independent variables of AoA and frequency. This analysis found a significant effect of AoA ($F(1,60) = 3.36$, $P < 0.1$) and frequency ($F(1,60) = 10.16$, $P < 0.05$). The interaction was not significant ($F(1,60) < 1$). A Cochran's C-test conducted on these data found no significant heterogeneity ($C(15,4) = 0.329$, $P = 0.62$). This still does not necessarily mean that the data are homogeneous but this statistic does show that a great deal of the heterogeneity has been removed by the log transformation. Further, the transformation produces data that better fit a normal curve (degree of skew = 0.206).

The problem of heterogeneity of variance is common to most studies employing reaction time data. Researchers often appeal to Ratcliff (1993) as a justification for not performing any transformations. While Ratcliff's work did demonstrate that a transformation is unlikely to affect the pattern of significance (as indeed did not happen in this case) this does not mean that we can make the same conclusions from a transformed analysis as we can from an untransformed analysis. If we were to accept the untransformed analysis then we must conclude that the two factors are additive. If we accept the transformed analysis, however, we conclude that the effects are interactive (this issue of the log transformation of ANOVA data is dealt with in Appendix A).

2.1.2. Discussion

The re-analysis has shown that there are several problems with using the Gerhand and Barry (1998) data to conclude additivity between AoA and frequency. First, there is the problem of the heterogeneous variance of the four cells when using the by-item data. This means that it may be dangerous to conclude too much from the analysis. Second, log-transforming the data makes them more homogeneous and the effects of AoA and frequency are still found. These significant factors now suggest that the effects may be multiplicative. These two possible alternative interpretations of the same data highlight some of the difficulties of using ANOVA to investigate two possibly related factors.

2.2. Multiple regression design: the Carroll and White (1973) data

Multiple regression designs are more numerous than factorial ANOVA designs in the literature on AoA effects; however, it is usually the case that the by-item data are

[2] By introducing an additive constant to the reaction time data it is possible to have a function that is asymptotic to a non-zero score. Such a procedure is consistent with the necessity that there is an absolute minimum time required for responding. Work on priming conducted by Lewis and Ellis (1999) found that the asymptote for decision reaction times was about 400 ms.

not provided and so many of these studies cannot be re-evaluated. Multiple regression experiments, like factorial ANOVA experiments, are not without their problems. One possible problem is that the independent measures employed (i.e. AoA and frequency estimates) are themselves random variables with their own intrinsic error of measurement. These errors may be greater or smaller for different values of the variables. In spite of this possible problem, multiple regression would appear to be the best method we have at present for investigating the effects of AoA and frequency on the speeds of a variety of tasks.

For a good example of the use of multiple regression to analyze AoA effects, it is useful to go back to the Carroll and White (1973) paper. This paper is where the laboratory-based research into AoA really began and, importantly, they provided the by-items data with which it is possible to re-evaluate their results. The method employed by Carroll and White was similar to many subsequent experiments exploring picture-naming: 50 participants were presented with a series of 94 line-drawn objects (plus nine practice items), their task being to name the object as quickly as possible during each presentation. Reaction times to the initial sound of a word were used as naming latencies. The words were each ascribed a value of word frequency (two different methods were employed), a measure of AoA based on ratings, and a measure of AoA based on objective data. The length of the word was also included in the analysis.

Although there was much to be commended about the hierarchical method of analysis employed by Carroll and White (1973), there are a number of issues with it: first, the frequency factor was log-transformed to form a standard frequency index (SFI) but the theoretical implications of this were not considered; second, the AoA was a non-linear rating scale with points on the scale being separated by different numbers of years; and third, and most important, the reaction time dependent factor was inverted for the analysis without the theoretical implications being considered. These theoretical implications of employing reciprocals are as follows.

The reciprocal analysis found a significant effect of AoA and an almost significant effect of frequency. Carroll and White (1973) examined the possibility of an interaction as they believed that only an interaction would mean that the AoA effects could be accounted for in terms of cumulative frequency. They found no interaction and so concluded that the effects could not be the result of cumulative frequency. When interpreting the results of Carroll and White (1973), however, it must be remembered that the reciprocal of reaction times was used (this point was ignored by Carroll and White themselves and led to their making inappropriate conclusions). This means that (assuming significant non-interacting effects of AoA and frequency) the regression model was

$$1/\mathrm{RT} = K + A\ln(\mathrm{freq}) + B(\mathrm{AoA}) \tag{4}$$

where A and B are constants, and, so putting Eq. (4) into terms of reaction times, we get

$$\mathrm{RT} = \frac{1}{(K + A\ln(\mathrm{freq}) + B(\mathrm{AoA}))} \tag{5}$$

which is not a linear combination of frequency and AoA. The analysis conducted, therefore, does not suggest additive effects but, rather, that the relationship between the two factors and reaction time is curvilinear and interactive.

2.2.1. Alternative linear and log-linear analyses

We employed three analyses to examine the effects of AoA and frequency on reaction time as summarized in Table 2. The first method was a reciprocal method as employed by Carroll and White (1973) themselves. The dependent variable was the inverse of the reaction times. There was a significant effect of the log-transformed frequency and a significant effect of the AoA (transformed from their non-linear rating scale). Inclusion of an interaction term did not improve the model and reduced the significance of the main effects.

Many experiments subsequent to that by Carroll and White (1973) that have explored AoA effects have employed linear multiple regression (as used by Barry et al., 1997, and others): that is, the raw reaction time scores have been used as the dependent variable rather than the reciprocal. For this reason, a re-analysis of Carroll and White's data was conducted using this second, linear, method and it revealed a significant effect of the log-transformed frequency and a significant effect of AoA. Inclusion of the interaction term did improve the fit of the model (i.e. the interaction was a significant predictor of reaction time). When the interaction term was included in the regression model, the main effects were no longer significant. In addition, inspection of the residuals of this analysis indicated a presence of a curvilinearity, which is indicative of an inappropriate analysis in much the same way as heterogeneous variance is for an ANOVA. Further, the distribution of the reaction times is skewed which also violates the assumptions of the analysis.

The cumulative-frequency hypothesis suggests a further method of analysis as

Table 2
Summary of the analyses conducted on the Carroll and White (1973) data[a]

Analysis	Reciprocal	Linear	Log-linear
Dependent variable	1/RT	RT	ln(RT)
Independent variables (no interactions)			
ln(frequency)	0.343*	− 0.330*	− 0.359*
AoA	− 0.493*	0.517*	
ln(Age − AoA)			− 0.512*
R^2	0.587	0.603	0.622
Independent variables (with interactions)			
ln(frequency)	0.265	− 0.021	− 0.358*
AoA	− 0.189	− 0.669	
ln(Age − AoA)			− 0.517*
Interaction	0.364	− 1.421*	− 0.005
R^2	0.589	0.632	0.622

[a] The figures show the standardized beta scores. *Significant at the 0.05 level.

described in the instance-based account above. This hypothesis predicts that the relationship between reaction time and the factors of frequency and AoA will be of the form shown in Eq. (3). Hence, the model is that the log of the reaction time should be predicted by the log of the frequency and the log of the time for which an item is known. This third method of analysis revealed a significant effect of the log-transformed frequency and a significant effect of the log-transform of time known (i.e. participant age minus AoA). The interaction was not significant and its inclusion did not affect the significance of the main factors. Inspection of the residuals found no curvilinear patterns.

2.2.2. Discussion

The conclusion drawn from the analysis by Carroll and White (1973) was that there were additive effects of frequency and AoA as indicated by a non-significant interaction term. However, as this analysis deals with the reciprocal of the reaction time then non-interacting factors do not imply that those factors are independent. Had Carroll and White (1973) performed an analysis on their raw data then they would have found an interaction such that it rendered the main effects no longer significant.

The linear analysis, that shows such a strong interaction in the Carroll and White (1973) data, is the more typical kind of analysis employed in subsequent regression designs (e.g. Morrison, Ellis, & Quinlan, 1992). The question must be asked as to why these studies have not found the interaction between AoA and frequency effects. The reason for this failure to find an interaction may be the same as that described for the factorial ANOVA designs: that is, owing to the curvilinear relationships between AoA, frequency and reaction times, the interaction effects may be much smaller than the main effects and so non-significant interactions have been taken to indicate the absence of an interaction.

The conclusion from the re-analyses of the Carroll and White (1973) data is that no matter how the results are analyzed, they suggest that the effects of AoA and frequency are multiplicative (or at least interactive). This conclusion is contrary to that drawn by Carroll and White and is consistent with the account of frequency and AoA effects in terms of cumulative frequency.

2.3. Re-analysis of large-scale databases

The re-analysis of the Carroll and White (1973) data is important as these were first used to reject the idea of AoA effects being merely cumulative-frequency effects. Their object-naming experiment, however, was based on a relatively small set of objects (just 94 items). Two, much more extensive studies, have been conducted into the lexical-decision task (Balota, Cortese, & Pilotti, 1999) and the word-naming task (Spieler & Balota, 1997). These studies originally did not explore AoA effects but they have been re-analyzed here using AoA and frequency norms.

2.3.1. The Balota et al. (1999) data

Balota et al. (1999) investigated the time it takes to make a lexical decision to

2906 words. In their original analysis of reaction times, they did not investigate the effect of AoA because they were not interested in this factor. As the mean reaction times (for 30 subjects) for this set of words are available on the internet, a re-analysis was possible to test the predictions of the cumulative-frequency hypothesis (as expressed in Eq. (3)).

The re-analysis employed a multiple regression design on the logarithm of the reaction time data (minus 400 ms). The independent variables were derived from the AoA ratings available from the MRC Psycholinguistic Database and frequency ratings available from the Celex database. The frequency independent variable was the logarithm of the Celex written frequency. The time known independent variable was the logarithm of the difference between the subjects' mean age and the AoA rating (from Gilhooly & Logie, 1982) converted into an age score. An interaction term was also included. Words were only included in the analysis if they were available in all the relevant databases. This left 553 words on which the analysis was conducted.

The results of the analysis found a significant effect of log of frequency ($b = -0.329$, $t(549) = 1.977$, $P < 0.05$), a significant effect of log of time known ($b = -0.721$, $t(549) = 5.934$, $P < 0.05$), and a non-significant effect of the interaction ($b = 0.042$, $t(549) = 1.368$, $P > 0.05$). Removing the interaction did not affect the significance of the results. The addition of a simple AoA effect to the regression did not significantly improve the fit of the model. Further, the residuals did not show any curvilinear patterns.

2.3.2. The Spieler and Balota (1997) data

In a similar study, Spieler and Balota (1997) investigated the time it tasks to read 2820 monosyllabic words. Again, AoA was not incorporated in the original analysis. These data can be re-analyzed using Celex frequency norms and the Gilhooly and Logie (1982) AoA norms. The data set was reduced to 553 items for which each kind of data was available. Such an analysis was conducted in order to test the revised cumulative-frequency hypothesis. The dependent variable was log of the by-item mean reaction times. The independent variables were log of time known and log of frequency.

The results found a non-significant interaction, which was removed from the regression. Consequently, there was a significant effect of log of frequency ($b = -0.020$, $t(550) = 2.414$, $P < 0.05$) and a significant effect of log of time known ($b = -0.474$, $t(550) = 6.313$, $P < 0.05$). The addition of a simple AoA effect to the regression did not significantly improve the fit of the model and the residuals did not show any curvilinear patterns.

2.3.3. Discussion

The lexical-decision task (Balota et al., 1999) and the word-naming task (Spieler & Balota, 1997) each follow the pattern of effects predicted by the cumulative-frequency hypothesis. It can be concluded, therefore, that the log of lexical-decision time is predicted by the two near-additive factors of log of frequency and log of time known. Similarly, the log of word-naming time is predicted by the two near-additive factors of log of frequency and log of time known. These results are wholly consistent with the idea that frequency and AoA effects are each a feature of the items' cumulative frequencies.

3. General discussion

The main conclusion from the re-analyses presented here is that there is little or no evidence that AoA effects are anything more than a feature of a cumulative-frequency effect. Previous work that had been used to support the idea that AoA and frequency effects are separable or independent factors is either dubious as a result of inappropriate analysis or has been interpreted in an indefensible manner. The data that have been re-evaluated here are entirely consistent with a cumulative-frequency account in which AoA and frequency of encounter are not different processes.

The idea that frequency and AoA each reflect aspects of the cumulative frequency of items has long been rejected and it has been assumed that the cumulative-frequency hypothesis is unable to account for any of the data found. It has long been concluded, from inappropriate statistical analysis, that the effects of AoA and frequency are non-interacting. Data such as the Gerhand and Barry (1998) cell means clearly display a parallelogram that is indicative of additive factors. It was demonstrated mathematically, however, that additive factors are not necessarily inconsistent with the revised version of the cumulative-frequency hypothesis. Further, it was shown that by using a transformation that corrects for both the skew and the heterogeneity of the variance, analysis of these data leads to the conclusion that AoA effects and frequency effects are multiplicative.

The original interpretation of the Carroll and White (1973) data also demonstrates an inappropriate conclusion. From their reciprocal analysis, they found additive effects of AoA and frequency. One can observe that additive effects on a reciprocal are not a linear combination of the two main factors and so do not mean additive effects on actual reaction time. Alternatively, had they conducted a linear analysis as have more contemporary researchers, then they would have observed that there is indeed a significant interaction between the factors on the raw reaction times. Of course, there are many studies, besides that of Carroll and White (1973), that have employed multiple regression techniques to study AoA and frequency effects. These tend to employ linear methods in their analyses but still do not find interaction effects (e.g. Ellis & Morrison, 1998). The revision to the cumulative-frequency hypothesis explains why, when a power law of learning is applied, additive effects appear as a consequence of interactive factors.

The re-analysis of the Balota et al. (1999) and the Spieler and Balota (1997) data presented here represents one of the largest studies into the effects of AoA and frequency. Analysis of these large data sets found that the best predictor of reaction time was the total number of occurrences of the words (i.e. frequency times time known). There was no significant effect of AoA once the effect of cumulative frequency had been accounted for. This result suggests, therefore, that it is possible that cumulative frequency accounts for the effects of frequency and AoA in lexical-decision and word-naming tasks.

The re-analyses reported here, employing log-transformed data, produced non-significant effects for the interaction terms. It could be argued, therefore, that we are making conclusions from null results. We are not, however, using the null result to reject any hypothesis (i.e. we are not making any strong conclusions). Indeed, we are not rejecting any hypothesis at all. The work here has been focused upon demonstrating how a hypothesis that has been long rejected is, in fact, consistent with the data. Our preference for the cumulative-frequency hypothesis is based unashamedly upon parsimony. The conclusions we draw are that there is no evidence, at the present time, to reject the idea that cumulative frequency leads to AoA effects and frequency effects.

While the re-analyses of the data, in themselves, are insufficient to suggest that AoA effects are solely caused by cumulative frequency, they do demonstrate quite clearly that the data are consistent with the cumulative-frequency hypothesis. This is particularly important because it was exactly these data that led to the original rejection of the cumulative-frequency hypothesis. We maintain that it is entirely possible that AoA has a specific effect beyond that of cumulative frequency but there appears to be no clear evidence for such an effect from within the existing literature. It is the case, therefore, that the most parsimonious explanation for AoA effects – the cumulative-frequency hypothesis – has been rejected prematurely.

Acknowledgements

We wish to thank two anonymous reviews for their comments on earlier drafts. Also, we wish to thank everyone who contributed data for our analysis either form published appendices or via the internet including D.A. Balota, J.B. Carroll, M.J. Cortese, M. Pilotti, D.H. Spieler and M.N. White. We thank C. Barry for helpful discussions.

Appendix A

Log-transforming reaction time data will rarely change the significance of the particular terms in a two-by-two ANOVA (Ratcliff, 1993) but it does change the conclusions we make from it. This point is illustrated here.

Assume we have an ANOVA on reaction times with significant main effects for the two factors (A and B) and no interaction. We would conclude therefore that

$$\text{Predicted reaction time} = f(\text{A}) + g(\text{B}) \tag{A1}$$

where f and g are some functions on the two factors. That is, the effects of A and B on reaction times are additive.

Now assume we have an ANOVA on the log-transformed reaction times with significant main effects for the two factors (A and B) and no interaction exactly. We would conclude that

$$\text{Predicted ln(reaction time)} = f'(\text{A}) + g'(\text{B}) \tag{A2}$$

where f' and g' are some functions on the two factors. To convert Eq. (A2) to the form of simple reaction times we must take the exponential of both sides. Therefore, we get

$$\text{Predicted reaction time} = e^{[f'(\text{A})+g'(\text{B})]} \tag{A3}$$

or

$$\text{Predicted reaction time} = e^{f'(\text{A})}e^{g'(\text{B})} \tag{A4}$$

which can be rewritten, without loss of generalizability as

$$\text{Predicted reaction time} = f''(\text{A})g''(\text{B}) \tag{A5}$$

The conclusion for this result is that the two factors, A and B, exert interactive effects on reaction times.

Eqs. (A1) and (A5) both come from analyses that are equivalent in terms of which factors are significant. The conclusions drawn from these significances, however, are very different.

References

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* (p. 44). Los Angeles, CA: Psychonomic Society.

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*, *50A*, 560–585.

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.

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

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture naming latency. *Quarterly Journal of Experimental Psychology*, *12*, 85–95.

Clarke, H. H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359.

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.

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.

Ellis, H. D., & Young, A. W. (1977). Age-of-acquisition and recognition of nouns presented in the left and right visual fields: a failed hypothesis. *Neuropsychologia*, *15*, 825–828.

Feyereisen, P., van der Borght, F., & Seron, X. (1988). The operativity effect in naming: a reanalysis. *Neuropsychologia*, *26*, 401–415.

Gazzaniga, M. S. (1970). *The bisected brain*. New York: Appleton-Century-Crofts.

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. (in press). Age of acquisition and the role of phonology in the lexical decision task. *Memory and Cognition.*

Gilhooly, K. J., & Gilhooly, M. L. M. (1979). Age-of-acquisition effects in lexical decision and episodic memory tasks. *Memory and Cognition*, *7*, 214–223.

Gilhooly, K. J., & Logie, R. H. (1982). Word age-of-acquisition and lexical decision-making. *Acta Psychologica*, *50*, 21–34.

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

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: the case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*, 185–207.

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

Kanji, G. K. (1993). *100 statistical tests*. London: Sage.

Lachman, R., Shaffer, J. P., & Hennrikus, D. (1974). Language and cognition: effects of stimulus codability, name-word frequency, and age-of-acquisition on lexical reaction time. *Journal of Verbal Learning and Verbal Behavior*, *13*, 613–625.

Lewis, M. B. (1999). Age of acquisition in face categorisation: is there an instance-based account? *Cognition*, *71*, B23–B39.

Lewis, M. B., & Ellis, H. D. (1999). Repeated repetition priming in face recognition. *Quarterly Journal of Experimental Psychology*, *52A*, 927–955.

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

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

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165–178.

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

Nickels, L. A., & Howard, D. (1995). Aphasic naming: what matters? *Neuropsychologia*, *33*, 1281–1303.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273–281.

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

Rochford, G., & Williams, M. (1962). Studies in the development and breakdown of the use of names: part 1, the relation between nominal dysphasia and the acquisition of vocabulary in childhood. *Journal of Neurology, Neurosurgery and Psychiatry*, *28*, 222–233.

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vandervart pictures. *Behaviour Research Methods, Instruments and Computers*, *28*, 515–536.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, *6*, 411–416.

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.