Computational Evidence That Frequency Trajectory Theory Does Not Oppose But Emerges From Age-of-Acquisition Theory

Martial Mermillod,[a,b] Patrick Bonin,[b,c] Alain Méot,[d] Ludovic Ferrand,[d] Michel Paindavoine[c]

[a]*Laboratoire de Psychologie et Neurocognition, UMR CNRS 5105, Université Pierre Mendès-France, Grenoble, France*
[b]*Institut Universitaire de France*
[c]*Université de Bourgogne and CNRS*
[d]*Clermont Université, Université Blaise Pascal, and CNRS*

Abstract

According to the age-of-acquisition hypothesis, words acquired early in life are processed faster and more accurately than words acquired later. Connectionist models have begun to explore the influence of the age/order of acquisition of items (and also their frequency of encounter). This study attempts to reconcile two different methodological and theoretical approaches (proposed by Lambon Ralph & Ehsan, 2006 and Zevin & Seidenberg, 2002) to age-limited learning effects. The current simulations extend the findings reported by Zevin and Seidenberg (2002) that have shown that frequency trajectories (FTs) have limited and specific effects on word-reading tasks. Using the methodological framework proposed by Lambon Ralph and Ehsan (2006), which makes it possible to compare word-reading and picture-naming tasks in connectionist networks, we were able to show that FT has a considerable influence on age-limited learning effects in a picture naming task. The findings show that when the input–output mappings are arbitrary (simulating picture naming tasks), the links formed by the network become entrenched as a result of early experience and that subsequent variations in frequency of exposure of the items have only a minor impact. In contrast, when the mappings between input-output are quasi-systematic or systematic (simulating word-reading tasks), the training of new items was generalized and resulted in the suppression of age-limited learning effects. At a theoretical level, we suggest that FT, which simultaneously takes account of time and the level of exposure across time, represents a more precise and modulated measure compared with the order of introduction of the items and may lead to innovative hypotheses in the field of age-limited learning effects.

Correspondence should be sent to Martial Mermillod, Laboratoire de Psychologie et NeuroCognition, Université Pierre Mendes France, BP 47, 38040 Grenoble Cedex 9, France E-mail: Martial.Mermillod@upmf-grenoble.fr

1. Background to the debate

An important issue in the field of psychology is to determine whether items (words, objects, faces, etc.) that are acquired early in life are processed faster and more accurately by adults than those that are acquired later in life, that is to say whether there is a late influence of early acquisitions. It seems plausible that the order of word acquisition is a factor that is directly responsible for the ease of processing these words, and indeed, this is the crucial tenet of the ''AoA hypothesis.'' This issue has fueled a debate in psycholinguistics. A large number of studies have provided convincing evidence of age-limited learning effects in lexical processing tasks (Johnston & Barry, 2006; Juhasz, 2005 for reviews) using age-of-acquisition (AoA) norms collected either from adult ratings (subjective AoA norms) or from children's performance (so-called objective AoA norms). Using subjective AoA norms, AoA effects have been found in a large variety of tasks (e.g., object, face, and action naming; word reading; and lexical decision) and in different populations (e.g., children, young and old adults, monolinguals and bilinguals, and aphasics). Recent attempts to manipulate AoA have revealed that this factor has a reliable influence on the learning of artificial patterns (Stewart & Ellis, 2008) as well as on the learning of a foreign vocabulary (Izura et al., 2011) in laboratory settings. However, despite the fact that robust AoA effects have been found in a wide variety of behavioral tasks (and simulations; see below), there is an ongoing debate as to whether the *order of acquisition* of the words is *per se* an important factor in determining the ease with which they are processed in both normal and impaired adults or whether the order of acquisition of the words is the result of several other embedded factors. Indeed, as far as the learning of the words in a language is concerned, factors other than the order in which the words and∕or concepts are encountered are obviously involved and are also certainly responsible for the speed and the accuracy of acquisition (with the result that certain words are acquired before others). Among these factors are (a) the frequency with which the words are encountered (e.g., during a certain period of life and, throughout one's entire lifetime) and (b) the kind of relationships, for example, systematic, quasi-systematic, and arbitrary, that exist between different types of codes (e.g., between phonological and orthographic codes, and between semantic codes and phonological codes). Some words are more frequent during certain periods of acquisition (e.g., ''*fairy*'' during childhood) than during others (e.g., ''*tax*'' during adulthood), and some other words retain their frequency of exposure during the lifespan (e.g., ''*house*'' is a high-frequency word and ''*platypus*'' a low-frequency word during both childhood and adulthood). Words that are frequently encountered are acquired earlier than those that are encountered less frequently (Bonin, Barry, Méot, & Chalard, 2004; Hazard, De Cara, & Chanquoy, 2007; Zevin & Seidenberg, 2002, 2004). However, as we will see, the question of whether words that have been frequently encountered during an early period of acquisition (whatever the evolution of their

frequency later in life) are easier to process later in life than words encountered less frequently depends on the kinds of relationships existing between the different types of codes. In alphabetic languages such as English or French, there are quasi-systematic relationships between sound units and orthographic units, whereas the relationships between semantic units and phonological (or orthographic) units are arbitrary. When quasi-systematic relationships are present, what is learned from certain items can be generalized to other items and, as a result, the processing of new items is easier than when no such generalization is possible, as is the case with arbitrary mappings (Lambon Ralph & Ehsan, 2006). In two behavioral studies, Bonin et al. (2004) and Zevin and Seidenberg (2004) have suggested that lexical processing varies as a function of both the frequency trajectory of the words (i.e., the frequency of exposure at various points during cognitive development) and the kinds of relationships that exist between semantics, phonology, and orthography. However, as the AoA of words (that is to say, adult-rated AoA as is frequently used in empirical studies) also has an impact on lexical processing, it is still necessary to determine whether frequency trajectory (FT) has an effect independent of that of AoA. Moreover, these authors have shown that FT effects also depend on the kinds of relationships that exist between semantics, phonology, and orthography. More precisely, Bonin et al. (2004) have revealed reliable age-limited learning effects in both oral and written naming latencies (where the relationships between object names and semantics are arbitrary), but not in word reading and in spelling-to-dictation latencies (in alphabetic languages such as French or English, the relationships between orthography and phonology are quasi-systematic). Although the measurement of the AoA of words (but also the objective frequency norms that are most appropriate to properly index the frequency of encounter of the words; see, for instance, Brysbaert et al., 2011) has been a topic of debate (Bonin, Méot, Mermillod, Ferrand, & Barry, 2009; Bonin et al., 2004), there is a general consensus that both the *order/age of acquisition of the words* and their *objective frequency of use* exert an influence in certain lexical tasks, and, more particularly, those tasks that rely on semantic-to-lexical mappings (e.g., object and face naming). As we shall see below, for many years, the range of empirical evidence in support of AoA effects exceeded the explanatory capabilities of the corresponding theoretical accounts. However, the theoretical account provided by Ellis and Lambon Ralph (2000) has proved to be very influential and fruitful in that it has shifted the focus from the collection of empirical demonstrations of age-limited learning effects in different tasks to the methodological and theoretical discussion of these effects. This study obviously adopts a similar approach. Indeed, several hypotheses have been put forward in the literature to account for age-limited learning effects, and it is not our intention to review these here (see Johnston & Barry, 2006). As stated above, one of the most fruitful and attractive accounts of age-limited learning effects is the connectionist account initially put forward by Ellis and Lambon Ralph (2000).

2. Age-limited learning effects in connectionist networks

Although distributed neural networks have long been used to address various issues in word recognition and spoken word production studies, they have also recently been used to

address the computational basis of age-limited learning effects (e.g., Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006; Monaghan & Ellis, 2010; Zevin & Seidenberg, 2002). In these connectionist models, lexical frequency is encoded in the strength of the connections between the different types of representations, which are involved in recognizing and producing words (Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989). As far as connectionist simulations of age-limited learning effects are concerned, the main innovation introduced by Ellis and Lambon Ralph (2000) in their simulations was that some patterns were used in training from the outset (''early'' patterns), whereas other (''late'') patterns were only introduced after the network had spent time learning the early ones. Using this manipulation, they showed that the synaptic weights of the system exhibited considerable plasticity for the early items. Conversely, late acquired items were more difficult to encode because these synaptic weights had already become ''entrenched'' during the coding of the first items. On the basis of this reduction of the plasticity of the synaptic weights with learning, the authors were the first to show that the order of introduction of the encounters determines the number of errors produced by the connectionist network at the end of training. More precisely, the main findings from their simulations included the following: (a) The benefits of early entry into training could not be accounted for in terms of differences in the frequency of training between the early and late sets. More specifically, when the late set was trained at a slightly higher frequency than the early set following its introduction to ensure that cumulative frequency was matched at the end of training, the system still performed better on the early than the late set. (b) High frequency of training brought about an additional benefit that was independent of the effect of order of entry. (c) Once an early item had been learned, its frequency of training could be reduced by a factor of 10 with no harmful effect on its representation. However, some forgetting was observed if items were removed from training altogether. And (d) items that only entered training very late had been trained at a very high frequency if they were to be learned effectively.

Zevin and Seidenberg (2002) have proposed an alternative theoretical framework to account for age-limited learning effects. In their theoretical approach, the AoA of the items does not constitute an independent variable per se but is instead *an outcome variable that is actually determined by other factors*. Given the significant link between AoA and cumulative frequencies, these authors suggested that the frequency trajectory of the encounters was an important factor. Zevin and Seidenberg (2002) found no frequency trajectory effects on network performance with background items. However, they found a weak but significant FT effect in simulation three under very specific circumstances, namely when background items were eliminated and ''critical'' items with very few neighbors were used. This suggests that the neighborhood of the encounter might influence AoA effects in artificial, and presumably also in biological and neural networks. Thus, whereas Ellis and Lambon Ralph (2000) suggested that the order of introduction of the items is a factor *determining* the performance of artificial or biological systems, Zevin and Seidenberg (2002) proposed that frequency trajectory better indexes this performance. In this view, AoA is actually an *outcome variable* and not an independent variable factor. This theoretical approach has been proposed for both connectionist networks (Zevin & Seidenberg, 2002) and for behavioral data

(Bonin et al., 2004; Zevin & Seidenberg, 2004). Consequently, Lambon Ralph and Ehsan (2006) decided to further investigate this hypothesis of differential AoA effects in connectionist networks by manipulating the input–output relationship *while retaining the theoretical framework* proposed by Ellis and Lambon Ralph (2000), which holds that the order of introduction of the items determines age-limited learning effects. Lambon Ralph and Ehsan's (2006) simulations showed that as the outputs became more predictable on the basis of the inputs, that is, as the mappings went from arbitrary through ''quasi-systematic'' to ''systematic,'' the effects of order of entry into training were reduced; that is to say, the cost of late acquisition became progressively less. In other words, they found that the order of introduction of the items had an effect with arbitrary mapping (each input vector is associated with a unique but arbitrary output vector), but not with componential mapping (when the input-output relationship is determined by a systematic association rule). The aim of this study was to establish a broad harmonization between Ellis and Lambon Ralph (2000), Lambon Ralph and Ehsan (2006), and Zevin and Seidenberg (2002).

3. Toward a unified theoretical framework

In this study, we introduce a proposal that attempts to reconcile studies on the order of introduction of the encounters, on the one hand, and studies of frequency trajectory, on the other. While the main body of the theory comes from Ellis and Lambon Ralph (2000), all the connectionist models of AoA, including that of Zevin and Seidenberg (2002), bear on this theory. However, Zevin and Seidenberg (2002) proposed FT as another way to *operationalize* age-limited learning effects. Surprisingly, this new proposal did not give rise to a large number of studies. Why? In our view, one reason might be that FT has long been perceived as a competing theory in the community. In our opinion, if we are to move beyond this situation, it has to be firmly established that (a) order of acquisition and FT rely on the same theoretical framework (related to synaptic entrenchment in biological or artificial neural networks); (b) the Zevin and Seidenberg (2002) study reported a very small age-limited learning effect in very specific situations because these authors used a *word reading* task, whereas Lambon Ralph and Ehsan (2006) reported that these effects occur essentially in *picture-naming* tasks, and (c) the order of acquisition is computationally equivalent to an extreme case of FT. Although this last point might seem difficult to understand a priori, it will be clarified below by examining how FT and the order of entry of the items have been operationalized in the different simulations performed in the AoA field (Ellis & Lambon Ralph, 2000; Zevin & Seidenberg, 2002) and in our current connectionist simulations described below.

On the basis of this theoretical framework, we suggest that FT provides a theoretical framework allowing a finer (because modulated) analysis of the empirical data. Another possible reason why FT has not given rise to many empirical studies is that the effect of FT was obtained in a very specific situation (involving very few critical items as we shall describe below) simulating word reading. However, following Lambon Ralph and Ehsan's (2006) work, it remains now clear that word reading may not be an ideal context for the observation

of AoA effects. We will therefore show that FT theory is able to simulate AoA effects with the same reliability as Lambon Ralph and Ehsan (2006) achieved in their picture-naming tasks (and even more precisely in the sense that we can modulate these AoA effects).

In this unified theoretical framework, although AoA is not considered an outcome variable[1] determined by FT as suggested by Zevin and Seidenberg (2002), the two independent variables are intrinsically related. It is important to stress that the approach we have followed is original because we explicitly take account of (a) the frequency trajectory of the items as a function of (b) their order of introduction into the learning situation as operationalized by Lambon Ralph and Ehsan (2006), and (c) the kinds of mappings that exist between the items (and especially arbitrary mappings which simulate picture-naming tasks, i.e., the most common situation in which age-limited learning effects emerge), as well as (d) the influence of the most frequently represented items in natural language, namely flat items (Monaghan & Ellis, 2010). To date, and including even the most recent AoA studies (e.g., Dent, Johnston, & Humphreys, 2008; Raman, 2006), the frequency trajectory concept has rarely been taken into account in the investigation (or discussion) of AoA effects, and in some cases it has simply been claimed to be somewhat equivalent to AoA (Levelt, Roelofs, & Meyer, 1999; Moore, 2003).

We differentiate the theoretical order of acquisition of the encounters, which is an objective independent variable, from the empirical measures of AoA (e.g., adult-rated AoA norms), which are much more difficult to encode and are sometimes a source of controversy (Bonin et al., 2009). More precisely, although there has been little discussion concerning ''theoretical order/AoA,'' that is to say the idea that there is an objectively identifiable moment/age at which the different words of a language are acquired, there has been considerable debate concerning how to operationalize the ''theoretical AoA'' for words. The measures used to operationalize the ''theoretical AoA'' for words, which are either subjective AoA norms collected from adults or so-called ''objective'' AoA norms (a misleading term, in our opinion) derived from children's performances, have attracted significant criticism because they have been given the status of independent variables (Bonin et al., 2009; Zevin & Seidenberg, 2004). This does not mean that theoretical AoA (and, of course, subjective AoA) is not influenced by other factors, such as frequency of exposure and memory resources. However, this is the case for virtually all psychological variables. For instance, in the same way, as in studies investigating the effect of memory on IQ, memory itself can be influenced by other factors such as attention or perceptual fluency. This does not mean that it is not possible to investigate the effect of memory on IQ. Therefore, in this study, we do not deny the status of theoretical AoA as independent variable. Instead, we show that FT is a more precise way to operationalize/measure this concept than, for instance, the subjective adult AoA norms.

Fig. 1A illustrates a theoretical but objective AoA as an independent variable. If we represent time along an axis as in Fig. 1A, each word is objectively learned during a specific period of time and this information represents an objective independent variable.

As mentioned above, the problem with AoA is that ''objective AoA,'' which is obviously an independent variable, is very difficult to encode. The debate between FT and AoA has emerged because some measurements of ''subjective AoA,'' that is, a mean of subjective
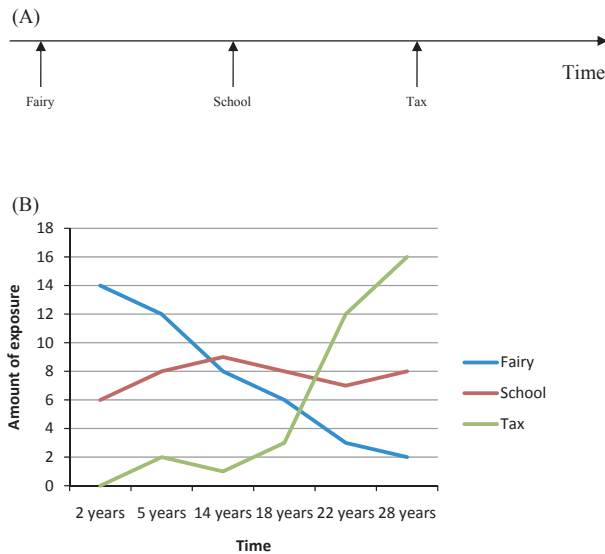
(A)

(B)

Fig. 1. (A) Objective age of acquisition encoded in the one-dimensional space provided by time. (B) Objective frequency trajectories encoded in the two-dimensional space provided by time and the amount of exposure to the items. This figure illustrates order of acquisition and frequency trajectories concepts, but it is not based on the real values of ''fairy,'' ''school,'' and ''tax'' words.

ratings corresponding to participants' estimates of the age at which they think they acquired words, are more similar to outcome variables than to independent variables (Bonin et al., 2009). Moreover, ''objective AoA'' as currently encoded in the literature suffers from similar problems (Bonin et al., 2009). Nonetheless, despite these technical difficulties in encoding the order of acquisition of the encounters, this does not mean that AoA does not exist as a real independent variable. Therefore, this study will show that FT (a) is an effective way of attenuating the problem of the encoding of the encounters and (b) is based on the same processes of synaptic entrenchment as order of acquisition but provides more information than order of acquisition. As shown in Fig. 1, order of acquisition is a unique, discrete, one-dimensional measure, whereas FT is an evolutive, continuous, two-dimensional measure. As a consequence, although AoA simply encodes the time at which a specific item is learned (Fig. 1A), FT encodes this information together with the amount of exposure to the items. It is therefore possible to encode both items of information in the two-dimensional space represented by time on the *X*-axis and the amount of exposure on the *Y*-axis (Fig. 1B).

The above explanation makes it clear that FT should not be viewed as opposed to AoA but instead as complementary to it. FT indicates the amount of exposure to the encounters at different periods of life. As illustrated in Fig. 1B, certain words are frequently encountered during childhood (''*fairy*'' and ''*giant*'') and less so in adulthood, while the opposite is true for other words (''*tax*'') and the frequency of exposure to yet other words remains stable over the life span; that is to say, they are equally rare (or frequent) in both childhood and adulthood. By means of connectionist simulations, we will show here that this theoretical framework leads to AoA effects that are qualitatively similar, while also permitting quantitative

modulation, to those obtained in simulations in which the order of entry alone is taken into account. We should observe that early acquired patterns are better encoded than late acquired patterns in the case of arbitrary but not systematic or quasi-systematic mappings. However, FT is not a one-dimensional, discrete variable (that can be specified by the time at which a specific word was learnt), but a two-dimensional, continuous variable defined by the frequency of the encounters at different periods of life. FT should therefore provide us with a better quantification of AoA effects than the order of introduction of the encounters. However, to ensure a fair comparison of the two factors, the methodological differences between Ellis and Lambon Ralph (2000) and Zevin and Seidenberg (2002) are reviewed to conduct the same computational analysis on the two types of independent variable within this theoretical framework.

4. Methodological and theoretical differences between FT and AoA computational studies

First, Ellis and Lambon Ralph (2000), Lambon Ralph and Ehsan (2006), Monaghan and Ellis (2010), and Zevin and Seidenberg (2002) did not use the same connectionist networks to test their hypotheses. Ellis and Lambon Ralph (2000) and Lambon Ralph and Ehsan (2006) used a standard back-propagation network. This algorithm is a very widespread and standardized algorithm used for a wide variety of cognitive processes such as language (Ellis & Lambon Ralph, 2000), as well as memory (French, 1999), visual recognition (Mermillod, Bonin, Mondillon, Alleysson, & Vermeulen, 2010; Mermillod, Vermeulen, Lundqvist, & Niedenthal, 2009), and semantic categorization (McClelland, McNaughton, & O'Reilly, 1995). The neural network used by Zevin and Seidenberg (2002) differed in some major specific respects from the standard back-propagation algorithm. For example, Zevin and Seidenberg (2002) included cleanup units (Hinton & Shallice, 1991) that mediate connections within the phonological unit layer. These feedback connections between the phonological and cleanup units create a dynamic system called an attractor network that settles into a stable pattern over time (Harm & Seidenberg, 1999). However, an attractor network, like the age-limited learning effects themselves, is intrinsically differentially affected by the ''sensitivity-stability'' dilemma (Hebb, 1949), namely the question of how to ensure that a memory model is simultaneously sensitive to, but not disrupted by, new inputs. Therefore, a direct comparison between Zevin and Seidenberg (2002) and Ellis and Lambon Ralph (2000) is not possible as Zevin and Seidenberg (2002) used cleanup units, whereas Ellis and Lambon Ralph (2000) did not.

Second, Ellis and Lambon Ralph (2000), Lambon Ralph and Ehsan (2006), and Zevin and Seidenberg (2002) used very different procedures to investigate the computational basis of AoA. One specific characteristic of Zevin and Seidenberg's (2002) simulations is that they included background items. Of the 2,891 items that constituted the training corpus, only 108 items were critical items whose frequency trajectories were manipulated. The remaining 2,783 background items improved the learning of the associative function linking input–output units. As mentioned by Monaghan and Ellis (2010): ''*A psychologically*

plausible developmental model of reading cannot include huge numbers of background items trained from the outset, as this runs quite contrary to the developing reader's experience'' (p. 12), Monaghan and Ellis (2010) argued that it was possible that the neural network generalized experience from the background items to critical items and that this process could therefore mask important AoA effects in their simulations. Although Zevin and Seidenberg (2002) used background items, both Ellis and Lambon Ralph (2000) and Lambon Ralph and Ehsan (2006) investigated the influence of age-limited learning effects on different types of input–output mapping, but without background items. It is worth stressing that AoA effects were obtained by Zevin and Seidenberg (2002) only when background items were removed from the simulations and for critical items having very few neighbors in quasi-systematic simulations (arbitrary relationships were not tested in their model). Therefore, an important methodological difference resulted from the use of what Ellis and Lambon Ralph (2000) and Lambon Ralph and Ehsan (2006) have defined as ''arbitrary mappings.'' Zevin and Seidenberg (2002) used ''critical'' items having few neighbors, namely words sharing similar phonological properties, to manipulate arbitrary mappings, whereas the simulations performed by Lambon Ralph and Ehsan (2006) involved completely arbitrary mappings (it should be remembered that such simulations are thought to approximate to picture naming, which involves arbitrary relationships between semantic codes and names). Thus far, and to our knowledge, there is no evidence that the small FT effect found by Zevin and Seidenberg (2002) in their modeling of word reading applies to the modeling of picture naming as proposed by Lambon Ralph and Ehsan (2006).

Third, we shall demonstrate that the findings reported by Lambon Ralph and Ehsan (2006), on the one hand, and Zevin and Seidenberg (2002), on the other, are theoretically different but not intrinsically opposed. We shall provide evidence that frequency trajectory can lead to similar predictions, and more important, are based on exactly the same properties of the neurally inspired learning mechanisms in parallel distributed processing (PDP) models. It is important to note that, in the following simulations, we have included items having a *stable frequency trajectory*. This point is far from trivial and represents an important baseline at the methodological level. Moreover, apart from the fact that these items can be taken as baseline items to improve the indexing of the true influence of items with decreasing or increasing frequency trajectories, they also represent a more ''natural case'' of what actually occurs in natural language. It is important to take account not only of the two types of frequency trajectories low-to-high frequency (''tax-like words'') and high-to-low frequency (''fairy-like words''). Real language contains only very few of these items, which represent a rather extreme case of the frequency trajectories exhibited by words in real language. As pointed out by Monaghan and Ellis (2010), most words that are early learned retain the relative frequency. For instance, these authors noted that the Educator's Word Frequency Guide (Zeno, 1995) lists 5,273 words which occur in the reading material suitable for Grade 1 readers with a frequency of 2 per million or more. Importantly, they reported a correlation of .812 between the frequency of those words in Grade 1 and their frequency in adult-level material (Grade 13 +). Thus, one of the important aspects of this study is the inclusion of stable frequency trajectories in the simulations. Due to this inclusion of items with stable trajectories, which represent the most frequent case of ''early acquired'' items, our

implementation of frequency trajectory does not reduce to an extreme case. Fourth, and directly related to the previous point, in our simulations we were careful to equalize cumulative frequency when investigating the FT effect. Finally, the interaction between FT and cumulative frequency was investigated.

We should emphasize that these differences constitute methodological differences that make it difficult (but not impossible) to perform certain comparisons among these studies of FT, on the one hand, and AoA on the other. However, there is a more important point that needs to be made at the theoretical level: Using a specific type of attractor network, Zevin and Seidenberg (2002) obtained a weak effect of FT for critical items that was limited to quasi-systematic mappings and occurred only when (a) background items were removed and (b) these critical items did not have neighbors. It is fair to say that these findings provide little support for a theory that sets out to improve on AoA theory. Why did the authors find such a small FT effect? To summarize, Zevin and Seidenberg (2002) provided evidence of a small effect of FT under very specific conditions because their simulation concentrated on word reading. However, Lambon Ralph and Ehsan (2006) reported that age-limited learning effects do not occur in word reading but do appear in picture-naming tasks. In other words, this study helps (a) show that FT is not a competing theory but an evolution of AoA theory (which was initially based on the order of acquisition of the items), and as such should be viewed as complementary to it, and it (b) provides computational evidence that FT gives rise to an age-limited learning effect in ''picture-naming-like tasks'' by investigating the effect of FT in a systematic and manner comparable to that adopted/initiated by Ellis and Lambon Ralph (2000).

5. Simulation 1: Frequency trajectory and AoA effects in artificial neural systems for arbitrary mappings

Simulation 1 tested the influence of frequency trajectories on arbitrary relationships between input and output units. The results of Lambon Ralph and Ehsan (2006), obtained within the theoretical framework of AoA, suggest that the nature of the relationships between input and output patterns is crucial for the emergence of age-limited learning effects on neural network performance. Thus, although age-limited learning effects should emerge in tasks requiring the involvement of arbitrary mappings, for example, in the case of face or object naming, few or no age-limited learning effects should be found in tasks requiring the involvement of systematic or quasi-systematic representations. Moreover, this pattern of findings has indeed been observed in the case of behavioral data (Bonin et al., 2004). In Lambon Ralph and Ehsan's (2006) simulations, an interaction between order of introduction and the lexical frequency of the patterns was observed. As far as the interaction between order of introduction and frequency of exposure is concerned, these simulations showed that the effect of order/AoA was greater for low-frequency than for high-frequency patterns (and this pattern was especially marked for the arbitrary patterns). Although this type of pattern is consistent with certain picture naming data (e.g., Barry, Morrison, & Ellis, 1997), an in-depth study has revealed that such interactions are rarely observed in the case

of picture naming latencies (Cuetos, Alvarez, Gonzales-Nosti, Méot, & Bonin, 2006). Although Lambon Ralph and Ehsan (2006) identified a significant effect of the order of introduction of the patterns when the relationships between input and output patterns were arbitrary, Zevin and Seidenberg (2002) did not find a reliable effect of frequency trajectory on network performance, when cumulative frequency was equalized across each training regime, and the overlap between the early and late training regime was eliminated (Simulation 4). According to these authors, what is learned from early items can be generalized to later items by means of associative learning functions inherent to the word reading task. However, as stated above, the general framework used by Zevin and Seidenberg (2002) is based on word reading tasks, whereas Lambon Ralph and Ehsan (2006) have shown that age-limited learning effects primarily occur in picture-naming tasks. In this study, we wish to show that frequency trajectories reveal similar and in the sense that they are modulated, more precise age-limited learning effects than the order of acquisition of the items in picture-naming tasks (using a similar methodology and similar material to those employed by Lambon Ralph & Ehsan, 2006). Moreover, we will show that these effects are based on the same neurally inspired processes as those responsible for the most recent results obtained, within the framework of AoA theory, by Lambon Ralph and Ehsan (2006) and to suggest that FT is, therefore, a more sophisticated form of the order of acquisition.

5.1. Material and procedure

The connectionist network was a standard three-layer back-propagation neural network. It was in all respects identical to the one used by Lambon Ralph and Ehsan (2006), namely a 100–50–100 neural network architecture. Like Ellis and Lambon Ralph (2000), and Lambon Ralph and Ehsan (2006), we did not include background items in the simulation relating to arbitrarily mapped items. For these items, the input and output vectors were 100 randomly generated binary vectors.

As can been seen from Fig. 2A, these 100 vectors were subdivided into five categories: (a) late acquired patterns, referred to as the ''late'' set, (b) decreasing FT patterns or the ''decreasing'' set (''fairy-like words''), (c) stable FT patterns or the ''stable'' set (''school-like words''), (d) increasing FT patterns or the ''increasing'' set (''tax-like words''), and (e) the ''early acquired'' patterns or the ''early'' set).

At time 1, the first 20 vectors (late acquired patterns) were encoded with a frequency of 1.1%, (each vector was presented once at each iteration of the neural network), the next 20 vectors (increasing FT) with a frequency of 11% (each vector presented 10 times at each iteration of the neural network), the next 20 vectors (stable FT) with a frequency of 22% (each vector presented 20 times at each iteration of the neural network), the next 20 vectors (decreasing FT) with a frequency of 33% (each vector presented 30 times at each iteration of the neural network), and the remaining 20 vectors (early acquired patterns) with a frequency of 33% (each vector presented 30 times at each iteration of the neural network). The first training period consisted of 1,000 epochs. Therefore, there were 91 presentations in each epoch and 1,000 epochs in each training period. During the second training stage of 1,000 epochs, early and late acquired patterns were encoded with a frequency of 1.1% (once

(A)

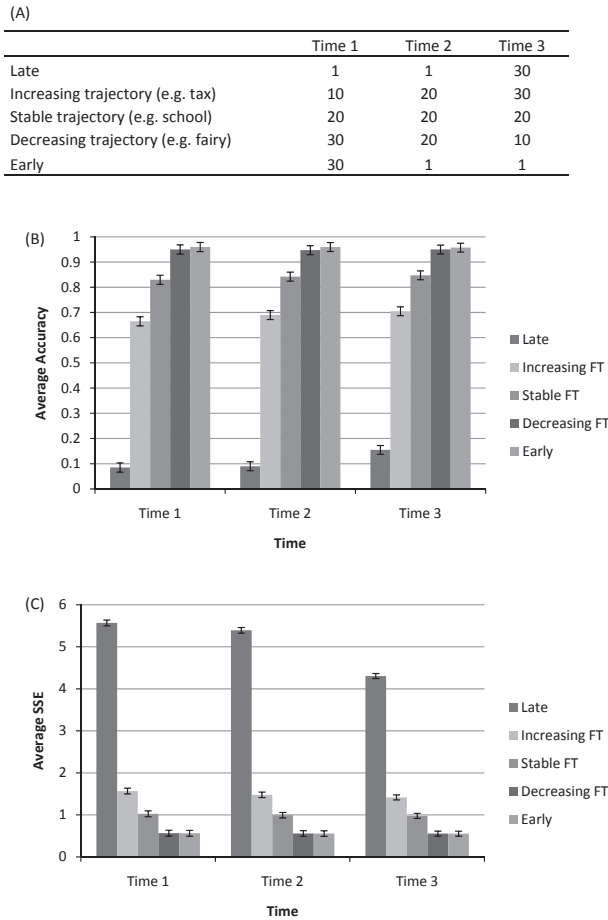|	Time 1	Time 2	Time 3
Late	1	1	30
Increasing trajectory (e.g. tax)	10	20	30
Stable trajectory (e.g. school)	20	20	20
Decreasing trajectory (e.g. fairy)	30	20	10
Early	30	1	1

Fig. 2. Frequency trajectories (A). Frequency trajectory effects on average accuracy (B) and Sum of Squared Errors (C) for arbitrary items.

at each iteration) and increasing, stable and decreasing FT with a frequency of 32% (each vector was presented 20 times at each iteration of the neural network). Finally, during the last training stage of 1,000 epochs, the first 20 vectors (late acquired patterns) were each encoded with a frequency of 33% (each vector was presented 30 times at each iteration of the neural network), the next 20 vectors (increasing FT) with a frequency of 33% (each vector presented 30 times at each iteration of the neural network), the next 20 vectors (stable FT) with a frequency of 22% (each vector presented 20 times at each iteration of the neural network), the next 20 vectors (decreasing FT) with a frequency of 11% (each vector presented 10 times at each iteration of the neural network), and the remaining 20 vectors (early acquired patterns) with a frequency of 1.1% (each vector was presented once at each iteration of the neural network). It should be noted that we presented early acquired patterns once at times 2 and 3 to prevent catastrophic forgetting. To compute an average accuracy

and the Sum of Squared Errors (SSE) of the network, we applied this training/test procedure for 20 runs. All occurrences are presented in Fig. 2A. To test the effect of FT independently of cumulative frequency, the cumulative number of occurrences for each FT was equated (60 occurrences for each FT condition). ''Early'' and ''Late'' patterns could not have the same cumulative frequency because they appeared only once in the training phase but were presented during the corresponding training phase with the same frequency as the ''high-to-low'' and ''low-to-high'' FT, respectively.

The underlying idea was to permit a direct comparison between FT and the order of introduction of the encounters to show that the order of acquisition, as operationalized in previous simulations (Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006), can be considered an extreme case of FT (i.e., as occupying a different point on the same continuum). We therefore included early and late acquired patterns to compare their influence with that of FT. We decided to use a slightly larger number of epochs (1,000 instead of 700 epochs) than in previous simulations (Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006) to ensure stable results within a more complex experimental design involving five training regimes and three points during development.

5.2. Results

We used the average accuracy (1-number of errors) and the SSE as the dependent variables (Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006; Zevin & Seidenberg, 2002). An output item produced by the neural network was defined as correct if the total discrepancy between the expected output and the observed output did not exceed 10%.

We conducted an ANOVA on the average accuracy rate and SSE with Training Period (Time 1, Time 2, and Time 3) x Type of (FT) set (Late, Increasing, Stable, Decreasing, and Early) as within-subject factor.

5.2.1. FT effects on accuracy

The statistical analysis revealed a significant effect of Type of set, $F(4, 76) = 411.6$, $MSE = 0.02$, $p < .001$, as well as a significant effect of the Training Period, $F(2, 38) = 7.8$, $MSE = 0.0004$, $p < .001$. The interaction between Type of set and Training period was significant, $F(8, 152) = 15.8$, $MSE = 0.0004$, $p < .001$, thus indicating that FT has a different effect depending on the training period (Fig. 2B).

Pairwise comparisons revealed that, at Time 1 (the ''childhood'' level), the late set produced a significantly lower accuracy than the increasing set, $F(1, 19) = 334.6$, $MSE = 0.01$, $p < .001$. Moreover, we also observed a significantly lower accuracy for the increasing set than for the stable set, $F(1, 19) = 25.2$, $MSE = 0.01$, $p < .001$, and a lower accuracy for the stable set than for the decreasing set, $F(1, 19) = 24.1$, $MSE = 0.006$, $p < .001$. However, we did not observe any significant difference between the decreasing set and the early set, $F < 1$. At Time 1, the neural network performance reflected only a frequency effect: The late set was presented only once and produced a lower level of accuracy than the increasing, stable, decreasing, and early sets presented 10, 20, 30, and 30 times, respectively. Thus, the first training period simply confirmed that the neural network was sensitive to the frequency

of the encounters. We obtained the same significant differences at Time 2 (the ''teenage'' level).

More interesting results were found at Time 3 (the ''adult'' level). Accuracy was significantly lower for the late set than for the increasing set, $F(1, 19) = 376.88$, $MSE = 0.008$, $p < .001$, and also significantly lower for the increasing set than the stable set, $F(1, 19) = 24.78$, $MSE = 0.008$, $p < .001$, and for the stable set than for the decreasing set, $F(1, 19) = 17.18$, $MSE = 0.006$, $p < .001$. The difference between the decreasing and early sets was not significant ($F < 1$). It should be noted here that the accuracy of the neural network was only 10% for late patterns. Such patterns constituted an extreme case with virtually no occurrences of the corresponding items in the first two training phases. This extreme case probably does not occur in real language. This helps to highlight the odd nature of the operationalization of the order of acquisition (which actually constitutes a very extreme case compared with frequency trajectories). However, it should also be noted that this result depends on other factors like the proportional frequency of late items compared with other items. It is possible to conjecture that a higher level of accuracy might be observed using the more realistic FT factor.

Taken together, the results show that the advantage of early acquired items observed in the early training stage persisted in the final training period despite a perfectly symmetrical inversion of the frequency of exposure.

5.2.2. FT effects on sum square error

As SSE represents a more precise variable, enabling us to measure the distance between expected outputs and observed outputs (Fig. 2C), we conducted the same ANOVA on SSE as had previously been performed for the average error rate. The ANOVA revealed a significant effect of Type of set, $F(4, 76) = 975.1$, $MSE = 0.22$, $p < .001$, as well as a significant effect of the Training period, $F(2, 38) = 433.1$, $MSE = 0.006$, $p < .001$. The interaction between Type of set and Training period was significant, $F(8, 152) = 291$, $MSE = 0.006$, $p < .001$.

More precisely, and as far as Time 1 is concerned, we observed a higher SSE on late acquired items than on increasing FT items, $F(1, 19) = 969.6$, $MSE = 0.16$, $p < .001$. Once again, at the end of Time 1, this difference simply reflected the fact that the neural network had had less exposure to ''late'' items than ''increasing'' items. We also observed a significantly higher SSE on ''increasing'' items than on ''stable'' items, $F(1, 19) = 41.2$, $MSE = 0.07$, $p < .001$, and a higher SSE for stable items than for decreasing items, $F(1, 19) = 38$, $MSE = 0.05$, $p < .001$. We did not observe any significant difference between decreasing items and early items ($F < 1$). However, this finding is not surprising given that both sets were presented with equal frequencies at Time 1.

At Time 2, the same significant differences were found. More interestingly, however, they were also observed at Time 3, thus showing that this effect of better training on the early set persisted. Pairwise comparisons revealed poorer training which resulted in a higher SSE on the late set than on the increasing set, $F(1, 19) = 927.2$, $MSE = 0.09$, $p < .001$, but also a significantly higher SSE on the increasing set than on the stable set, $F(1, 19) = 35.1$, $MSE = 0.05$, $p < .001$, and a higher SSE for the stable set than the decreasing set, $F(1,$

19) = 30.4, MSE = 0.06, $p < .001$. As in the first and second training period, we did not observe a significant difference between the decreasing and early sets, $F < 1$.

5.2.3. FT effects across time

This analysis examined the effect of the different types of FT across time (from Time 1 to Time 3). Concerning the late set, we observed a significant improvement in the effect of learning on SSE over time, $F(1, 19) = 386.04$, $MSE = 0.04$, $p < .001$. We observed the same effect on the increasing set, $F(1, 19) = 42.89$, $MSE = 0.005$, $p < .001$ and stable set, $F(1, 19) = 15.03$, $MSE = 0.001$, $p < .01$. More surprisingly, we also observed a significant reduction of SSE for the decreasing set, $F(1, 19) = 4.73$, $MSE = 0.002$, $p < .05$ despite the reduced exposure to the items. We even observed a marginally significant reduction of SSE for the early set, $F(1, 19) = 4.1$, $MSE = 0.002$, $p = .057$, despite the drastic reduction in the level of exposure.

5.3. Discussion of Simulation 1

An effect of the frequency trajectory variable was found on the neural network performance when the mappings between the input and output units were arbitrary. Compared with previous connectionist data, this finding means that it is possible to obtain an age-limited learning effect on arbitrary items without necessarily having to refer to the order of introduction of the encounters (Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006). Instead, this simulation shows that similar effects can be obtained through a manipulation of the frequency trajectories of the items, as suggested by Zevin and Seidenberg (2002). Nonetheless, Zevin and Seidenberg (2002) did not test the effect of FT on arbitrary relationships simulating picture naming. To our knowledge, this is the first computational evidence that FT is able to produce AoA effects in picture naming-like tasks but also that FT can be operationalized as a continuous variation in the order of acquisition. This effect is consistent with AoA effects observed in behavioral picture naming data (Bonin et al., 2004) when cumulative frequency is controlled for. A key aspect of the data from Simulation 1 is that the training frequencies at Time 1 dominated the outcome, not only at the end of this period but also at the end of the periods corresponding to Times 2 and 3. At all three times, the best performance was observed on the Early and Decreasing sets, followed by Stable and then Increasing, with Late being the worst set by quite some margin, even after Time 3 when the frequency of the items had risen from 1 to 30. Thus, it remains clear that once the network structure has been formed through early experience, later changes have relatively little effect, a phenomenon that is usually referred to as ''entrenchment.'' Also, once an item has become established through early experience, little subsequent exposure is required to maintain its representation (see also Ellis & Lambon Ralph, 2000; Simulation 12). In line with previous data (Ellis & Lambon Ralph, 2000; Munro, 1986; Zevin & Seidenberg, 2002), these findings suggest that age-limited effects arise from a generic aspect of learning; that is to say, the plasticity of the network decreases with learning. We shall return to the issue of plasticity in the General Discussion. The consequence of the reduction of network plasticity is that the point during learning at which items are first encountered has a long-term and

stable effect on behavioral data. This is also confirmed by the current analysis of the different FT across time. These findings clearly show a strong entrenchment effect that was powerful enough to reduce SSE over time even despite the drastic reduction in the level of exposure for the early acquired items.

6. Simulation 2: Frequency trajectory effects in artificial neural systems for quasi-systematic (2A) and systematic (2B) mappings

Two further simulations were run using new patterns of vectors having (a) a quasi-systematic relationship between the input and output layers (Simulation 2A) and (b) systematic input–output relationships (Simulation 2B). As in the previous simulation, we used the input and output patterns provided by Lambon Ralph and Ehsan (2006). The frequency trajectory of the items was manipulated, while their cumulative frequency was held constant. These data instantiate the quasi-regular mapping of English or French. This context is thought to operationalize and simulate reading aloud (or spelling-to-dictation) in alphabetic languages. In the light of the findings reported by Zevin and Seidenberg (2002), no reliable effect of frequency trajectory was predicted on neural network performance with a quasi-systematic and a systematic coding of the input–output relationship (except in one very specific condition, namely critical items without background items). This represents a clear contrast to Lambon Ralph and Ehsan's (2006) study, which obtained small but significant age-limited learning effects for quasi-systematic relationships in a simulation that used the order of introduction of the items as independent variable. In Simulation 2B, we expected the regularities of the input–output patterns to completely suppress age-limited learning effects in the artificial neural network.

6.1. Material and procedure

The network was identical to the one used in Simulation 1. For the quasi-systematic items (Simulation 2A), we used the structure relationship provided by Lambon Ralph and Ehsan (2006). The quasi-regular mappings were created by dividing the 100 unit vectors into three sections (33, 34, and 33) to represent a CVC-like word. We used the identical abstract patterns for 10 consonant and 10 vowel components generated by Lambon Ralph and Ehsan (2006) to produce a hundred representations that were formed by joining the CVC patterns using a Latin-square type combination. In other words, each input vector $C_n V_n C_n$ was associated with an output vector $C_n V_n C_{n+1}$. Likewise, all the 10 consonant and vowel patterns were used 10 times each in both the onset and offset positions. The frequency trajectories of the encounters were identical to Simulation 1.

Turning to Simulation 2B, in the same way as for the quasi-systematic data, 100 unit vectors were created to form CVC-like words based on the 10 consonant and 10 vowel components generated by Lambon Ralph and Ehsan (2006). The only difference compared with Simulation 2A was that each input vector $C_n V_n C_n$ was associated with itself as an output vector. In other words, the connectionist network was an auto-associator neural network that

permitted the reproduction of perfectly predictable input–output correspondences (as in Turkish; see Raman, 2006). Frequency trajectories were strictly identical to Simulation 1.

6.2. Results

As far as accuracy is concerned, the quasi-systematic and systematic relationships produced a clear ceiling effect resulting in 100% correct responses for each FT and early or late AoA items, thus clearly indicating that this new associative procedure was easier than that used in Simulation 1 (as reported by Lambon Ralph & Ehsan, 2006 on the basis of the order of acquisition of the items). We therefore ran an ANOVA on SSE, which enabled us to obtain a more precise analysis of the results (because SSE is a continuous variable, which makes it possible to describe very small differences).

We conducted an ANOVA on SSE (Fig. 3) with Training period (Time 1, Time 2, and Time 3) x Type of sets (Late, Increasing, Stable, Decreasing, and Early) as within-subject variable on 20 runs of the training/test procedure (equivalent to 20 different subjects).
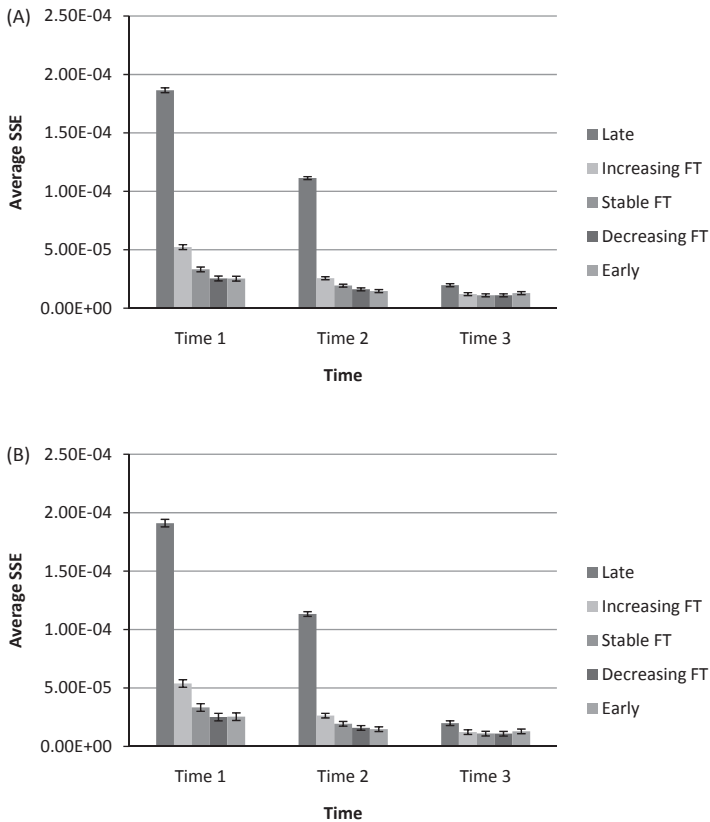

Fig. 3. Frequency trajectory effect on average Sum of Squared Errors for systematic (A) and quasi-systematic items (B).

6.3. Simulation 2A: Quasi-systematic items

The statistical analysis revealed a significant effect of Type of sets, $F(4, 76) = 476.8$, $MSE = 0.0001$, $p < .001$, as well as a significant effect of the Training period, $F(2, 38) = 1353.96$, $MSE = 0.0001$, $p < .001$. The interaction between Type of sets and Training period was significant, $F(8, 152) = 503.7$, $MSE = 0.0001$, $p < .001$.

More precisely, and as can be seen from Fig. 3A, pairwise comparisons revealed a significant, albeit very weak, effect of frequency at the end of Time 1. The late set produced higher SSE than the increasing set, $F(1, 19) = 413.2$, $MSE = 0.0001$, $p < .001$, and we observed a significant effect of the increasing set compared with the stable set, $F(1, 19) = 596.4$, $MSE = 0.0001$, $p < .001$, as well as of the stable compared with the decreasing set, $F(1, 19) = 299.8$, $MSE = 0.0001$, $p < .001$. This effect was not significant for the decreasing compared with the early sets. At the end of the training regime (Time 3), we observed a drastic reduction of the SSE for all the types of sets. However, albeit very small, the difference between late and increasing sets was significant, $F(1, 19) = 303.1$, $MSE = 0.0001$, $p < .001$, as was the difference between the increasing and stable sets, $F(1, 19) = 38.7$, $MSE = 0.0001$, $p < .001$. However, we have to note the striking fact that, although we observed a frequency effect at the end of Time 1, the quasi-systematic relationship between the input and output units produced only very tiny SSE differences between the different training conditions, and especially at the end of the training regime (Time 3). In other words, such small differences could probably not be transposed to the behavioral level. There was a drastic reduction in SSE on quasi-systematic items compared with that observed for arbitrary items, $F(1, 291) = 621.37$, $MSE = 99.07$, $p < .001$.

6.3.1. FT effects across time

Concerning the late set, we observed a significant effect of improvement of learning on SSE as exposure to this set increased over time, $F(1, 19) = 646.9$, $MSE = 0.001$, $p < .001$. We observed the same effect on increasing FT, $F(1, 19) = 3257.9$, $MSE = 0.001$, $p < .001$ and stable FT, $F(1, 19) = 7620.3$, $MSE = 0.001$, $p < .001$. As for arbitrary items, we also observed a significant reduction of SSE for decreasing FT, $F(1, 19) = 6000.3$, $MSE = 0.001$, $p < .001$, despite the reduction of exposure to these items, as well as a significant reduction of SSE on early acquired items, $F(1, 19) = 2538.4$, $MSE = 0.001$, $p < .001$, again despite the drastic reduction in the level of exposure to these items.

6.4. Simulation 2B: Systematic items

We observed a significant effect of Type of sets, $F(4, 76) = 1147.3$, $MSE = 0.0001$, $p < .001$, a significant effect of the Training period, $F(2, 38) = 3293.9$, $MSE = 0.0001$, $p < .001$, and a significant interaction between Type of sets and Training period, $F(8, 152) = 1112.7$, $MSE = 0.0001$, $p < .001$. As in the case of quasi-systematic relationships, pairwise comparisons for Time 1 revealed that the late set produced higher SSE than the increasing set, $F(1, 19) = 920.2$, $MSE = 0.0001$, $p < .001$, a significant effect of increasing set compared with the stable set, $F(1, 19) = 332.4$, $MSE = 0.0001$, $p < .001$, and also a

significant effect of the stable set compared with the decreasing set, $F(1, 19) = 139.2$, $MSE = 0.0001$, $p < .001$. The difference between decreasing and early sets was not significant ($F < 1$). We also observed a substantial reduction of the SSE at the end of Time 3. Again, even though very small, the difference between the late and increasing sets was significant, $F(1, 19) = 751.7$, $MSE = 0.0001$, $p < .001$, as was the difference between the increasing and stable sets, $F(1, 19) = 12.4$, $MSE = 0.0001$, $p < .01$.

6.4.1. FT effects across time

As for the quasi-systematic relationships, we again observed a significant effect of improvement of learning on SSE as exposure to the late set increased over time, $F(1, 19) = 1489.1$, $MSE = 0.001$, $p < .001$. The same findings were obtained for the increasing set, $F(1, 19) = 3,887.2$, $MSE = 0.001$, $p < .001$ and the stable set, $F(1, 19) = 6,256.8$, $MSE = 0.001$, $p < .001$. We also observed a significant reduction of SSE for the decreasing set, $F(1, 19) = 8,539.5$, $MSE = 0.001$, $p < .001$, despite the reduction of exposure to these items, as well as a significant reduction of SSE on the early set, $F(1, 19) = 3,598$, $MSE = 0.001$, $p < .001$. Once again, this improvement occurred despite the drastic reduction in the level of exposure to these latter two sets (Fig. 3B).

6.5. Discussion of Simulations 2A and 2B

To summarize the findings from Simulations 2A and 2B, the network performance on items having quasi-systematic or systematic mappings was much better than when it was given the task of learning items having arbitrary mappings (Simulation 1). Learning continued to take place across Times 2 and 3, and all the sets had been learned well by the end of Time 3. Even the Late set was well learned after Time 3 after its frequency had increased from 1 to 30. The findings obtained for quasi-systematic and systematic mappings are consistent with previous behavioral (Bonin et al., 2004; Zevin & Seidenberg, 2004) and computational studies (Lambon Ralph & Ehsan, 2006; Zevin & Seidenberg, 2002), thus showing that no age-limited learning effects emerge when reading aloud (or spelling-to-dictation) in alphabetic languages in which grapheme-to-phoneme (or phoneme-to-grapheme) correspondences are perfectly predictable. Early in training, the network performance is better on items that are trained more often; that is to say, a frequency effect occurs during the initial phase of the training regime. At a computational level, we found a significant but tiny difference only at the level of SSE, thus resulting in a higher error rate for the late acquired or the increasing FT items. However, it should be noted that we did not find any AoA effect on accuracy scores when input–output mapping was systematic or quasi-systematic. These results make it possible to reconcile the previous data reported by Lambon Ralph and Ehsan (2006) with those obtained by Zevin and Seidenberg (2002). Our data indicate that age-limited learning effects are drastically reduced by quasi-systematic or systematic relationships, even though an examination of the small SSE errors still makes it possible to observe this type of tiny AoA effect. Moreover, it is important to point out that, although we observed a small difference on SSE, there was virtually no effect on accuracy. These results mean that the effect is very small at the computational level (as observed by

Zevin & Seidenberg, 2002) and is therefore likely to be very difficult to obtain at a behavioral level (Bonin et al., 2004). The findings obtained in Simulations 2A and 2B are compatible with the hypothesis that age-limited learning effects are difficult to obtain when the mappings between input and output units are quasi-systematic (or systematic) as has been empirically observed in word reading in alphabetic languages such as French (Bonin et al., 2004), English (Zevin & Seidenberg, 2004), or Italian (Burani, Arduino, & Barca, 2007). At a computational level, the findings from Simulation 2 suggest that the phenomenon of plasticity loss previously observed by Munro (1986) might be considerably reduced in componential representations when cumulative frequency is adequately controlled for. Furthermore, the above simulations show that the theoretical framework provided by FT is able to simulate results similar to those obtained by Lambon Ralph and Ehsan (2006) on the basis of order of acquisition.

Taken together, the findings suggest that frequency trajectories are able to simulate the effect of order of acquisition, while at the same time permitting a better quantification of age-limited learning effects than the simple order of introduction of the encounters. Our findings strongly suggest that variations in frequency over time have more of an impact in networks that process more systematic mappings. However, the frequencies of encounter of the items can decline over time without affecting the quality of the representations that are formed by the network, as evidenced by the performance observed on the Early items in Simulations 2A and 2B.

As the findings obtained for systematic and quasi-systematic relationships are similar and age-limited learning effects were obtained only when arbitrary mappings were used, we decided, in a final simulation, to focus on the interaction between frequency trajectory and cumulative frequency using items having *arbitrary* mappings.

7. Simulation 3: Interaction of frequency trajectory and cumulative frequency for arbitrary mappings in connectionist networks

Lambon Ralph and Ehsan (2006) have shown that the order of introduction of the items and cumulative frequency interact, with the result that the effect of order of introduction on network performance is greater for low-frequency than for high-frequency items. Moreover, these authors have also shown that the interaction between the frequency of the patterns and the order of introduction is stronger, when the mappings between input and output patterns are arbitrary than when they are systematic or quasi-systematic. Indeed, they found that order of introduction/AoA and frequency interacted overall, but that this interaction was only reliable in the arbitrary mapping simulation.

As far as the behavioral data are concerned, as claimed by Lambon Ralph and Ehsan (2006), there is a surprising paucity of empirical data concerning the issue of whether AoA and word frequency interact. The reason may be related to the close relationship between rated/objective AoA norms and word frequency measures, which makes it difficult to use a factorial design. In effect, it is quite difficult to find high-frequency stimuli that are acquired late in life and low-frequency stimuli that are acquired early because most words acquired

early are high in frequency and this frequency remains high (Monaghan & Ellis, 2010). As far as we know, the interaction between word frequency and AoA was first tested by Barry et al. (1997) in a multiple regression analysis on spoken naming times. These authors found a reliable interaction, with a larger frequency effect being obtained on late acquired than on early acquired items. However, this interaction was not obtained using objective AoA norms, thus suggesting that the interaction may not be robust. Meschyan and Hernandez (2002) performed a factorial crossing of the AoA and frequency variables, as well as the delay (0 or 2200 ms) between picture onset and a naming cue, but did not find a significant interaction between the two variables in immediate naming (i.e., interval 0). In the Lambon Ralph and Ehsan (2006) study, even though an interaction between word frequency and AoA was found to affect both picture naming and word reading latencies, the interaction between word frequency and AoA was not reliable on items. In addition, Chalard, Bonin, Méot, Boyer, and Fayol (2003), who used different measures of word frequency, did not find a reliable interaction between the two factors in picture naming latencies in French. Finally, Cuetos et al. (2006) conducted an extensive study to test the interaction between word frequency and AoA in picture naming latencies and concluded that the modulation of the word frequency effect as a function of the age/order of acquisition of the words is not a robust finding in picture naming performance in adults. To summarize, the behavioral results are controversial given that some of them indicate an interaction between AoA and cumulative frequency (Barry et al., 1997; Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006), where others do not (Chalard et al., 2003; Cuetos et al., 2006; Meschyan & Hernandez, 2002). The next simulation investigated this potential interaction within the theoretical framework of frequency trajectories.

7.1. Material and procedure

The connectionist network was the same three-layer back-propagation neural network used in Simulations 1 and 2. The input and output vectors were the same 100 binary vectors generated in Simulation 1. For reasons of clarity, we removed the early and late acquired sets as Simulation 1 showed that these items behave in a similar fashion to the decreasing and increasing sets. The first 33 vectors were encoded with a frequency of 16.7% (each vector was presented once at each epoch), the next 34 with a frequency of 33.3% (each vector was presented twice at each epoch), and the remaining 33 with a frequency of 50% during the first step of training (each vector was presented three times at each epoch). During the second and third steps of the training, the frequency trajectory manipulation was identical to that of Simulations 1 and 2. To obtain two levels of cumulative frequency, high-frequency (HF) vectors, 1–16, 34–50, and 68–83 were presented to the neural network twice as frequently as the low-frequency (LF) vectors, with the result that each HF vector was presented twice, four times or six times at each epoch, respectively. In other words, we used the same FT and input/output mapping as in Simulation 1 for LF items and simply multiplied the LF frequencies by two to obtain the HF frequencies. As in the previous simulations, this training regime was used for 20 runs. A summary of the cumulative frequency and frequency trajectory manipulations is provided in Fig. 4A.

(A)

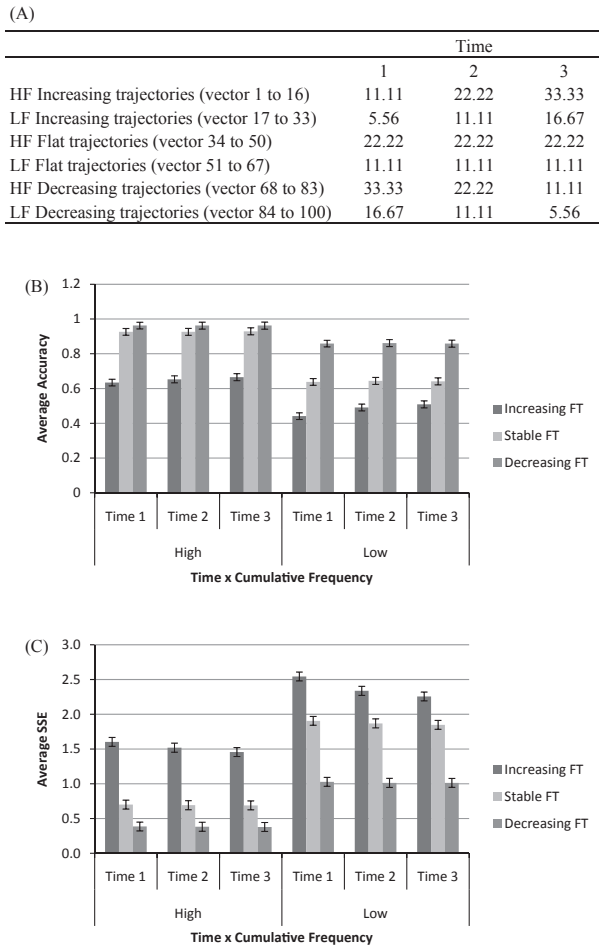|	Time		
	1	2	3
HF Increasing trajectories (vector 1 to 16)	11.11	22.22	33.33
LF Increasing trajectories (vector 17 to 33)	5.56	11.11	16.67
HF Flat trajectories (vector 34 to 50)	22.22	22.22	22.22
LF Flat trajectories (vector 51 to 67)	11.11	11.11	11.11
HF Decreasing trajectories (vector 68 to 83)	33.33	22.22	11.11
LF Decreasing trajectories (vector 84 to 100)	16.67	11.11	5.56

Fig. 4. Frequency trajectories and cumulative frequencies (A). Frequency trajectory and cumulative frequency effect on average accuracy (B) and Sum of Squared Errors (C) for arbitrary items.

7.2. Results of Simulation 3

We conducted an ANOVA on SSE and average accuracy with Training Period (Time 1, Time 2, and Time 3), Frequency Trajectory (Increasing, Stable, and Decreasing), and Cumulative Frequency (High and Low) as within-subject factors.

7.2.1. FT effects on accuracy

The statistical analysis revealed a significant effect of FT, $F(2, 38) = 200.8$, $MSE = 0.02$, $p < .001$, a significant effect of cumulative frequency $F(1, 19) = 116.6$, $MSE = 0.03$, $p < .001$, and a significant effect of the training period, $F(2, 38) = 26$, $MSE = 0.0004$, $p < .001$. Interestingly, the interaction between FT and cumulative frequency was significant,

$F(2, 38) = 9.22$, $MSE = 0.03$, $p < .001$, meaning that FT has a different effect depending on the cumulative frequency of the encounters.

As in Simulation 1, at the end of Time 1, accuracy was significantly higher for the decreasing low-frequency set (mean accuracy = 0.86) than for the stable low-frequency set (mean accuracy = 0.64; $F(1, 19) = 53.6$, $MSE = 0.01$, $p < .001$), and the error rate significantly lower for the stable low-frequency set than for the increasing low-frequency set (mean accuracy = 0.44; $F(1, 19) = 32.2$, $MSE = 0.01$, $p < .001$). Turning to the high-frequency items, we observed a lower error rate for the stable set (mean accuracy = 0.93) than for either the increasing set (mean accuracy = 0.63; $F(1, 19) = 141.3$, $MSE = 0.006$, $p < .001$) or the decreasing frequency sets (mean accuracy = 0.96; $F(1, 19) = 5.5$, $MSE = 0.002$, $p < .05$). We obtained the same significant differences at Time 2.

The most important results were those found at Time 3. We observed a significantly higher accuracy for the decreasing low-frequency set (mean accuracy = 0.85) than for the stable low-frequency set (mean accuracy = 0.64; $F(1, 19) = 49.9$, $MSE = 0.009$, $p < .001$). The observed error rate was significantly lower for the stable low-frequency set than for the increasing low-frequency set (mean accuracy = 0.51; $F(1, 19) = 13.6$, $MSE = 0.01$, $p < .01$). As far as the high-frequency set is concerned, we observed a lower error rate for the stable set (mean accuracy = 0.93) than for either the increasing set (mean accuracy = 0.66; $F(1, 19) = 101.5$, $MSE = 0.007$, $p < .001$) or for the decreasing frequency set (mean accuracy = 0.96; $F(1, 19) = 4.98$, $MSE = 0.002$, $p < .05$). However, for both high- and low-frequency items, the increasing set was correctly recognized by the neural network significantly less often than the decreasing set, $F(1, 19) = 143.5$, $MSE = 0.006$, $p < .001$ and $F(1, 19) = 104.3$, $MSE = 0.011$, $p < .001$, respectively.

7.2.2. FT effects on SSE

The main effect of cumulative frequency was significant, $F(1, 19) = 213.1$, $MSE = 0.33$, $p < .001$, with high-frequency items leading to lower SSE than low-frequency items (Fig. 4B). This effect is similar to the main frequency effect reported by Lambon Ralph and Ehsan (2006). The main effect of frequency trajectory was also significant, $F(2, 38) = 221.8$, $MSE = 0.21$, $p < .001$, as was the effect of the Training period, $F(2, 38) = 148.1$, $MSE = 0.002$, $p < .001$. Moreover, the interaction between cumulative frequency and frequency trajectory at the end of the training period (Time 3) was significant, $F(2, 38) = 8.5$, $MSE = 0.26$, $p < .001$.

At the end of Time 1, we observed a significantly lower SSE for the low-frequency decreasing set (mean SSE = 1.03) than for the low-frequency stable set (mean SSE = 1.9; $F(1, 19) = 40.1$, $MSE = 0.1$, $p < .001$), and a significantly lower SSE was observed for the low-frequency stable set than for the low-frequency increasing set (mean SSE = 2.54; $F(1, 19) = 58.1$, $MSE = 0.13$, $p < .001$). Turning to the high-frequency sets, the results were the same as for the low-frequency sets at the end of Time 1, namely a lower SSE was observed for the decreasing (mean SSE = 0.38) than for the stable set (mean SSE = 0.7; $F(1, 19) = 175.4$, $MSE = 0.04$, $p < .001$), where a lower SSE was observed for the stable set than for the increasing set (mean SSE = 1.6; $F(1, 19) = 32.5$, $MSE = 0.03$, $p < .001$).

These differences persisted during both the second and the final training session. At Time 3, a significantly lower SSE was found for the low-frequency decreasing set (mean SSE = 1.01) than for the low-frequency stable set (mean SSE = 1.84, $F(1, 19)$ = 62, MSE = 0.11, $p < .001$). The same difference was also observed between the stable and increasing sets (mean SSE = 2.26; $F(1, 19)$ = 15.5, MSE = 0.1, $p < .001$). For the high-frequency set, the SSE was lower on the decreasing (mean SSE = 0.38) than the stable items (mean SSE = 0.69; $F(1, 19)$ = 32.04, MSE = 0.03, $p < .001$), and the same significant difference was observed between the stable and increasing sets (mean SSE = 1.46; $F(1, 19)$ = 119.2, MSE = 0.05, $p < .001$). The increasing set produced significantly higher SSE than the decreasing set in the case of both high-frequency, $F(1, 19)$ = 263, MSE = 0.04, $p < .001$ and low-frequency items, $F(1, 19)$ = 123, MSE = 0.12, $p < .001$.

7.2.3. FT effects across time

As for arbitrary items, we again tested the effect of FT over time in this new simulation but, in this case, for different cumulative frequencies. Concerning high-frequency words, we observed a significant reduction of SSE from Time 1 to Time 3 for the increasing set, $F(1, 19)$ = 128.1, MSE = .002, $p < .001$, as well as for the stable set, $F(1, 19)$ = 5.57, MSE = .003, $p < .05$. This difference did not reach significance for the decreasing set ($p = .14$). With regard to the low-frequency words, we observed a similar significant reduction of SSE over time for the increasing set, $F(1, 19)$ = 89.5, MSE = .009, $p < .001$ and the stable set, $F(1, 19)$ = 12.8, MSE = .002, $p < .01$, and, in this case, also for the decreasing set, $F(1, 19)$ = 4.5, MSE = .0005, $p < .05$.

7.3. Discussion of Simulation 3

The findings from Simulation 3 can be summarized as follows. With high training frequencies, the network became entrenched after Time 1. Performance on the decreasing and stable sets was good after Time 1 and these sets remained well learned after the following two Times 2 and 3. In contrast, the increasing set (the late set in this simulation) benefitted very little from the increase in frequency of occurrence across Times 2 and 3. As far as the lower frequencies are concerned, the decreasing set was again well-learned at Time 1 and this learning persisted over the next two periods (Times 2 and 3). The stable set was rather less well learned at Time 1 and did not change much across Times 2 and 3. The level of learning on the increasing set was only marginal between the end of Time 1 through to the end of Time 3. As was the case for Simulations 1 and 2, the analysis of FT effects over time revealed a significant reduction in errors on both the increasing and stable, as well as on the decreasing set. Thus, the effect of FT seems to be greater for LF items than for HF items (because a type of ceiling effect occurs for HF words).

Focusing on frequency trajectory effects, both high-frequency and low-frequency items showed age-limited learning effects. However, as reported by Lambon Ralph and Ehsan (2006) for AoA, the FT effect was more specific to low-frequency than high-frequency items, thus resulting in a significant interaction between FT and cumulative frequency. Our current findings suggest that this interaction is probably due to an overtraining on

high-frequency items, which results in an absence of any difference between the AoA effects on accuracy induced by stable FT, on the one hand, and decreasing FT, on the other. These computational results lead to new behavioral hypotheses, and we might expect a similar absence of difference between stable and decreasing FT in humans. Moreover, we have to note that the interaction, albeit significant, was very weak. This might explain the controversial results previously reported in the literature (Barry et al., 1997; Chalard et al., 2003; Cuetos et al., 2006; Lambon Ralph & Ehsan, 2006; Meschyan & Hernandez, 2002).

8. General discussion

Our starting point was that any theory of lexical processing in adults has to account not only for the factors which determine the speed and accuracy of lexical processing but also the reasons why these factors are thought to be influential. This issue has been debated at length in the psycholinguistic literature, in particular with regard to the influence of the frequency with which words are encountered and their age of acquisition. As the measures corresponding to these two factors are correlated, there has been some controversy about whether the two factors have a genuine influence (e.g., Barry et al., 1997). A very large number of studies have reported an influence of the age of acquisition of words in a wide variety of lexical tasks (Johnston & Barry, 2006; Juhasz, 2005 for reviews). As a result, some researchers have claimed that AoA has a universal influence in lexical processing (e.g., Raman, 2006). The key assumption behind the AoA hypothesis is that the order in which words are acquired has a direct influence on lexical processing speed and accuracy in mature cognitive systems. Therefore, based on the extensive modeling undertaken by Ellis and Lambon Ralph (2000), Lambon Ralph and Ehsan (2006), Monaghan and Ellis (2010), and Zevin and Seidenberg (2002), connectionist simulations have been used to provide evidence for an explicit theoretical account of why the order of acquisition of the items per se has a long-lasting influence.

Following previous studies (Bonin et al., 2009; Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006; Zevin & Seidenberg, 2002, 2004), a connectionist theory has been put forward to account for age-limited learning effects in mature cognitive systems. Within this connectionist framework, AoA is explained by the reduction of synaptic plasticity in previously trained artificial neural systems. This theory makes explicit statements regarding the influence of AoA, cumulative frequency, and frequency trajectory in lexical processing. According to this theory, the frequency trajectory of items is an important underlying factor determining the age/order of acquisition of words (Bonin et al., 2004, 2009; Zevin & Seidenberg, 2002), as well as the speed and accuracy of processing of items. This is especially true in the case of tasks that require the use of arbitrary mappings, such as object and face naming tasks which demand the mobilization of semantic-to-lexical mappings. Connectionist data suggest that, when input/output relationship are arbitrary, the more often and earlier an individual is exposed to a word, the better the corresponding representations formed by the cognitive system will be. The basic idea underlying FT theory is therefore to replace a discrete, one-dimensional coding of the AoA variable by a

continuous, two-dimensional variable, which takes account of the amount of exposure to encounters over time. Frequency trajectory can thus be used to investigate age-limited learning effects in lexical processing (from word reading to object/face-naming tasks). According to this theory, the influence of frequency trajectory on a mature cognitive system is confined to the specific cases where learning about certain items cannot be generalized to new items (when specific links between input and output patterns have to be learned). As suggested by Zevin and Seidenberg (2002), when generalization is possible, as in the case of items having systematic or quasi-systematic mappings (i.e., the links involved in word reading or spelling-to-dictation in alphabetic languages), the frequency trajectory of the items has relatively little impact on the mature performance of the network. However, it is worth remembering that Zevin and Seidenberg (2002) did not investigate the effect of FT on arbitrary relationships (the links involved in object or face naming) and that their findings therefore provide little evidence of the effect of FT as arbitrary relations have been shown to be the most sensitive to AoA effects (Lambon Ralph & Ehsan, 2006). This might account for why FT has not been investigated more enthusiastically in subsequent empirical studies (Bonin et al., 2009). However, in accordance with the theoretical framework proposed by Zevin and Seidenberg (2002), our simulations make clear (a) that when items with arbitrary mappings have to be learned, these items become entrenched as a result of early learning and later learning exerts only little influence, thus resulting in clear AoA effects, but also (b) that FT and order or acquisition (as operationalized in Lambon Ralph & Ehsan, 2006) can be considered different points along one and the same continuum. Below, we shall address the more fundamental issues relating to the meaning of age-of-acquisition effects. Importantly, this theory has also been confirmed by a small number of behavioral studies (Bonin et al., 2004; Izura et al., 2011; Stewart & Ellis, 2008; Zevin & Seidenberg, 2004) and it is to be hoped that the current theoretical and computational support for the use of FT will encourage more extensive research within this theoretical framework. At a behavioral level, this raises the question of the status of subjective AoA measures that might actually constitute a performance variable which has to be accounted for. The computational findings reported here clearly show that FT is a reliable candidate that provides more informative data, making it possible to account for the AoA effect observed in psychological data.

   At a behavioral level, Zevin and Seidenberg (2002) have suggested that the age of acquisition of the items is an outcome variable to be accounted for rather than a factor which itself has an influence on the ease with which the patterns are learned. Items that are encountered more frequently during an earlier period of acquisition are learned first (Bonin et al., 2004; Hazard, De Cara, & Chanquoy, 2008; Zevin & Seidenberg, 2002) and therefore determine the age of acquisition of the items, as currently recorded at the behavioral level. It will be necessary to resolve this debate at the behavioral level (Bonin et al., 2009). At a theoretical level, our aim here was to assist in the development of a unified theoretical framework by drawing on the methodological and theoretical strengths of both approaches. We are fully aware of the importance and the theoretical existence of the order of acquisition of items (the time at which a specific word is encoded in the cognitive system), even though this measure has proved to be difficult to record in humans and has given rise to some debate

(Bonin et al., 2004, 2009). However, FT theory represents a step forward in that it provides a more precise way to address the question of age-limited learning effects by taking account of the continuous amount of exposure resulting from the encounters. Similarly, this theory makes it possible to address the effects on the behavior of a mature network that is still learning by making use of the same principles as previously applied, namely (a) when the items are learned, (b) how often the network is exposed to them, and (c) as a function of the types of mapping between these items. These different aspects have not previously been addressed within one and the same study.

Following the original formulations of the AoA hypothesis—which did not include the concept of frequency trajectory—and the subsequent connectionist implementation by Lambon Ralph and Ehsan (2006)—which did not make any explicit reference to it—Lambon Ralph and Ehsan (2006) reported that the type of mapping between different representations (arbitrary vs. componential) interacted with the order of introduction of the patterns in the training session. More precisely, these simulations showed that patterns introduced early in training and complemented by later patterns which were trained alongside them in a cumulative and interleaved manner were recognized better than late patterns introduced at the end of learning. In this study, we investigated the possibility that FT theory is able to account for these differential age-limited learning effects. As in Lambon Ralph and Ehsan (2006), the type of relationships between input and output units was also manipulated (Simulations 1, 2, and 3) in order to show that frequency trajectory is a useful additional parameter that can help explain the results obtained when the order of introduction of the encounters is manipulated. One aspect of this study worth stressing is that we included items with a stable frequency trajectory. As shown by Zevin and Seidenberg (2002), the inclusion of this baseline is important because (a) it permits us to gain a more accurate understanding of the true influence of decreasing or increasing frequency trajectories on network performance, and more importantly, and (b) the stable trajectory represents a more natural case of the words that are encountered early when learning a language. As reviewed in the Introduction, the correlation between child and adult frequencies in American-English is very high, meaning that the words that children are exposed to at an early age retain their frequency later in life. Finally, in Simulation 3, the combined influence of frequency trajectory and cumulative frequency in the presence of arbitrary mappings was investigated. A simple three-layer feedforward model—the same as the one used by both Ellis and Lambon Ralph (2000) and Lambon Ralph and Ehsan (2006)—was trained on sets of patterns with different frequency trajectories. As had already been shown by Lambon Ralph and Ehsan (2006) on the basis of the *order of acquisition* of the items, the age-limited learning effects provided by FT varied as a function of the type of mapping. Simulations 1, 2, and, 3 revealed that frequency trajectory had a reliable influence on AoA effects, when arbitrary mappings, but not quasi-systematic or systematic mappings, were used. Therefore, age-limited influences were found to be related to the type of learning that takes place between different kinds of representations. This study makes it clear that when the mappings between input and output units are not predictable, or in other words when the relationships between codes are arbitrary, the links that are formed by the network become entrenched as a result of early experience. Later experience and variations in the word frequencies involved in the learning regime do

little to change them. The learning in our network was clearly dominated by the phenomena of entrenchment and loss of plasticity (see below for a more extensive discussion), irrespective of the high or low cumulative frequency of the items. Thus, when the items to be learned are unpredictable, individual links have to be memorized and age-limited learning effects are expected to have a long-lasting influence. In contrast, when the items to be learned are more predictable (or clearly predictable) and the relationships are systematic or quasi-systematic; that is to say, when generalization is possible, variations in the frequency of exposure to them over time have an impact on networks. However, even in this case, frequencies can decline over time without affecting quality of representation (as shown by the performance on the early items in Simulations 2A and 2B).

8.1. Implications and future directions for research

Over the last 10 years, a number of simulations based on the AoA hypothesis have been run in an attempt to account for age-of-acquisition effects in terms of age-dependent reductions in plasticity (Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006; Zevin & Seidenberg, 2002). Ellis and Lambon Ralph (2000) have suggested that there is an explicit link between age-limited learning effects and neural network plasticity. We fully agree with the hypothesis that a reduction in neural plasticity can account for age-limited learning effects. Indeed, we have shown, on the basis of the current connectionist simulations, that the same neurally inspired processes constitute the theoretical underpinning of FT theory and have demonstrated, we believe for the first time, that FT could produce reliable and stable AoA effects in simulations of ''picture naming-like'' tasks. Moreover, we have shown that FT permits a modulation of the responses of the neural network, which is not present if only the order of introduction of the encounters is considered. This comparison between AoA and FT raises a more fundamental question about AoA. As shown in these simulations, we can modulate the *order of acquisition* in terms of ''early'' and ''late'' patterns as Ellis and Lambon Ralph (2000) and Lambon Ralph and Ehsan (2006) have done and show that these situations represent a formal equivalent to extreme cases of FT. However, in this computational framework, it is more difficult to define the notion of *age of acquisition*. Connectionist simulations clearly show that the acquisition of items in an artificial neural network (as well as in infants) is a continuous, and not a discrete, process. It is possible to determine the period during which we were exposed to a word, but it is more difficult to pinpoint when it was *acquired*. In this study, for example, we had to decide what threshold to adopt to decide if the input–output associations for a specific word were sufficiently reliable for it to be considered ''acquired.'' However, one key question that has to be answered is what does it mean to define a word as ''acquired'' or ''not acquired?'' Is a single exposure to a word sufficient to define the word as acquired? Is a single production of a word sufficient to define the word as acquired? Does the acquisition of a word require a certain level of accuracy in its production? If such a level is required, we are entitled to ask what that level is and what dimensions are involved (semantic representations, phonological/orthographic representations, all of these)? We assume here that the earlier debate about whether AoA measures constitute an independent or a dependent variable may, in part, be

related to this problem and suggest that FT might constitute a more precise (because modulated) and more objective independent variable. We therefore strongly recommend recording this measure to construct databases that will be useful for future psycholinguistic research.

A more specific question that may be raised is *why is there a reduction in neural plasticity* in the systems that we are studying? Neural plasticity can be modulated in artificial neural networks to produce strong a ''entrenchment'' effect at one extreme of the continuum (i.e., when the neural network has very low plasticity/high stability) and catastrophic forgetting at the other end (i.e., when the neural network has very high plasticity/low stability).[2] This question will have to carefully addressed in future research in neural modeling. It should be remembered that in the Ellis and Lambon Ralph (2000) simulations, catastrophic interference was observed on the early items when these were entirely replaced by the later items during learning by the network. Could the reduction in neural plasticity be a consequence of the particular frequency that we used? That is to say, if had we used frequency trajectories other than those employed here, for example, with each word appearing in only one time block but neither before nor after it, might we have observed very different learning outcomes? If this scenario, which is based on Ellis and Lambon Ralph's (2000) findings, were indeed the case, then a reasonable prediction would be that later acquired items erase earlier acquired items, as can be seen in the catastrophic interference phenomenon (French, 1999 and below). Although this is clearly an interesting issue, it is very rare, when learning words in a given language, to find that the words to which we have been exposed over a period of time suddenly cease to occur in the language in question (except for a small number of words which become completely obsolete). However, one interesting natural social situation is that of adopted children who have learned a language for a period of their lives and been exposed to the words it contains, and then cease to be exposed to these words when they move to a different country. Pallier et al. (2003) attempted to reveal native language traces in eight adult Koreans who had been adopted during childhood (between 3 and 8 years of age) by French families. These individuals had been completely cut off from and not reexposed to Korean since their arrival in France (15–20 years prior to testing). At the subjective level, these adults claimed to have no knowledge of their mother tongue. The performances in different linguistic tasks (sentence identification in Korean versus different languages, word recognition, and fragment detection) of the Korean adoptees and a control group consisting of native French speakers were compared and no reliable differences were found between the groups. Functional magnetic resonance imaging (fMRI) was also used to monitor brain activations while the adults were listening to sentences in French, Korean, and two other unknown languages. None of the adoptees exhibited activation specifically in response to Korean compared with the unknown languages. Moreover, their pattern of activation in response to French sentences was quite similar to that of the native French speakers. Thus, it would seem that early exposure to a language is not enough to leave permanent traces in the brain (see also Ventureyra, Pallier, & Yoo, 2004). Re-exposure to the language is necessary to maintain the stored representation, as otherwise ''naturalistic'' catastrophic interference is observed. These behavioral findings fit nicely with what has been observed at a computational level by Ellis and Lambon Ralph (2000). These authors examined the case of

items acquired early by the connectionist network before suddenly ceasing to be presented to the network, as in the case of certain real words that appear primarily in nursery rhymes or children's stories. Except among adults who themselves look after children, the frequency of such words will be lower in adulthood than it was in early childhood. The simulations performed by Ellis and Lambon Ralph (2000) showed that once an early set of patterns has been well learned by the network, a frequency of presentation that is greatly reduced from the original level is enough to maintain the quality of the representations. It was only when the early patterns ceased to be trained at all that representations in the studied network suffered and catastrophic interference was observed on these items. It should be noted that the conditions that give rise to catastrophic interference in neural networks (see French, 1999 for further discussion) rarely occur in learning in humans because in real life (French, 1999), the learning experiences consist of items that are interleaved, as in the learning of objects and their names, letters, and their sounds, and so on. As shown by Hetherington and Seidenberg (1989), relearning occurs even with neural networks that exhibit catastrophic interference, a finding which indicates that initial or early acquisitions are not completely erased. Thus, a prediction in humans that was recently confirmed by Bowers, Mattys, and Gage (2009) is that individuals who cease to be exposed to a language they knew in their childhood are able to (re-)learn this language more easily than individuals who have never learned the language in question.

   Is reduced plasticity a feature of biological systems? As far as language acquisition is concerned, the reduction in plasticity reflects the idea that there is a critical period of age during which exposure to language must take place in order for language to be acquired normally. This is referred to as the critical period hypothesis (Seidenberg & Zevin, 2006). This period is thought to be followed by subsequent restrictions of the ability to learn in the form of a loss of plasticity. However, evidence for such restrictions to language learning capabilities after a putative critical period remains controversial, even though it seems that language learning is indeed age dependent. What could be responsible for this loss of plasticity in learning? At the biological level, there are candidates for intrinsic changes to neural networks that may limit plasticity: synaptic pruning, changes in the number and distribution of neurotransmitter receptors, and the maturation of inhibition. Importantly, simulation studies (including ours) strongly suggest that *learning itself* plays a role in the reduction of plasticity. However, apart from the computational evidence, other evidence in favor of this hypothesis takes the form of song learning in zebra finches (Zevin, Seidenberg, & Bottjer, 2004). Zebra finches typically learn song during a sensitive period that closes early in adulthood after which new song elements are not added and existing elements are not lost. However, they exhibit plasticity beyond the end of the sensitive period that can be extended by altering the bird's experience. Zevin et al. (2004) went on to show that when white noise was used to prevent adult birds from hearing their own songs (without damaging their hearing), there was a negative impact on the birds' songs. This evidence is entirely consistent with the computational approach and indicates that continued exposure to the bird's own song is necessary for the maintenance of the bird's capabilities. However, when the noise was discontinued, even though the birds were not able to learn from exposure to a tutor (which suggests a limitation to plasticity), song did change over

time in response to auditory feedback, thus suggesting a continued capacity to learn. According to Seidenberg and Zevin (2006), the acquisition and gradual entrenchment of representations that support stereotyped song become increasingly resistant to change as a result of the process of learning itself. Thus, the reduction of plasticity due to learning in neurally inspired networks is clearly compatible with evidence revealed by learning in biological systems. Finally, as far as language is concerned, the loss of plasticity is not negative in nature as the knowledge that is acquired is established in a way that allows generalization. On the contrary, this loss can be thought of as a positive adaptive feature. Importantly, and as pointed out explicitly by Seidenberg and Zevin (2006), when unpredictable facts, such as the correspondence between objects and their names, are learned, that is, in cases where age-limited learning effects are most frequently observed, these effects differ from those associated with critical periods in two respects. First of all, age-limited learning effects concern particular items rather than systematic aspects of knowledge and, second, the conditions that give rise to these effects do not lead to the failure to acquire new items. This contrasts with the case of critical periods, which result in the inability to acquire, generalizable, systematic knowledge. According to these authors, and in contrast to the standard view, *the process of learning creates neurobiological changes that reduce plasticity*. Whether a critical period truly exists for language acquisition, the loss of plasticity is a phenomenon that should not to be equated with critical periods.

To conclude, this study represents a valuable contribution because it provides clear computational support for the idea that frequency trajectory is a valuable way of addressing the issue of age-limited learning effects due to the fact that frequency trajectory makes it possible to take account explicitly of both the order of introduction of the items during learning and their exposure levels over the entire acquisition process. It should therefore permit a more precise modulation of AoA effects than order of acquisition in future studies involving naming tasks.

Acknowledgments

Notes

1. It can be argued that not only AoA but also word frequency is an ''output variable,'' which has effects because it ''stands for'' other factors. As we have argued in Bonin et al. (2009), the empirical measures of word AoAs, that is, ''objective-'' or subjective-rated AoA norms, are characterized as a behavioral outcomes because the way they are

measured depends directly on the participants' performance (naming accuracy in children and ratings in adults), whereas word frequency is derived from the analyses of corpora: It is the number of times a word is found in a corpus. It might also be thought that the order of acquisition of the words and their frequency of use could exert more direct effects on lexical processing, irrespective of the factors that cause variations in AoA and frequency.

2. We thank an anonymous reviewer for having brought this to our attention.

References

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*, *50*, 560–585.

Bonin, P., Barry, C., Méot, A., & Chalard, M. (2004). The influence of age of acquisition in word reading and other tasks: A never ending story? *Journal of Memory and Language*, *50*, 456–476.

Bonin, P., Méot, A., Mermillod, M., Ferrand, L., & Barry, C. (2009). The effects of age of acquisition and frequency trajectory on object naming. *Quarterly Journal of Experimental Psychology*, *62*, 1–9.

Bowers, J. S., Mattys, S. L., & Gage, S. H. (2009). Preserved implicit knowledge of a forgotten childhood language. *Psychological Science*, *20*, 1064–1069.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, *58*, 412–424.

Burani, C., Arduino, L. S., & Barca, L. (2007). Frequency, not age of acquisition affects Italian word naming. *European Journal of Cognitive Psychology*, *19*, 828–866.

Chalard, M., Bonin, P., Méot, A., Boyer, B., & Fayol, M. (2003). Objective age-of-acquisition (AoA) norms for a set of 230 object names in French: Relationships with other variables used in psycholinguistic experiments, the English data from Morrison et al.(1997) and naming latencies. *European Journal of Cognitive Psychology*, *15*, 209–245.

Cuetos, F., Alvarez, B., Gonzales-Nosti, M., Méot, A., & Bonin, P. (2006). Determinants of lexical access in speech production: Role of word frequency and age of acquisition. *Memory & Cognition*, *34*, 999–1010.

Dent, K., Johnston, R. A., & Humphreys, G. W. (2008). Age of acquisition and word frequency effects in picture naming: A dual-task investigation. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *34*, 282–301.

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.

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, *3* (4), 128–135.

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

Hazard, M.-C., De Cara, B., & Chanquoy, L. (2007). Normes d'âge d'acquisition objectif des mots et recherche de prédicteurs : Importance du choix de la base de fréquence lexicale. *L'Année Psychologique*, *107*, 427–457.

Hazard, M.-C., De Cara, B., & Chanquoy, L. (2008). Normes de fréquence cumulée et trajectoire fréquentielle comme prédicteurs de l'âge d'acquisition. *L'Année Psychologique*, *108*, 487–516.

Hebb, D. O. (1949). *Organization of behavior*. New York: Wiley & Sons.

Hetherington, P., & Seidenberg, M. (1989). ''Is there 'catastrophic interference' in connectionist networks?'' *Proceedings of the 11th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

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

Izura, C., Pérez, M. A., Agallou, E., Wright, V. C., Marín, J., Stadthagen-Gonzalez, H., & Ellis, A. W. (2011). Age/order of acquisition effects and the cumulative learning of foreign words: A word training study. *Journal of Memory and Language*, *64*, 32–58.

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition*, *13*, 789–845.

Juhasz, B. (2005). Age-of-Acquisition effects in word and picture identification. *Psychological Bulletin*, *131*, 684–712.

Lambon Ralph, M. A., & Ehsan, S. (2006). Age of acquisition effects depend on the mapping between representations and the frequency of occurrence: Empirical and computational evidence. *Visual Cognition*, *13*, 884–910.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–437.

Mermillod, M., Bonin, P., Mondillon, L., Alleysson, D., & Vermeulen, N. (2010). Coarse scales are sufficient for efficient categorization of emotional facial expressions: evidence from neural computation. *Neurocomputing*, *73*, 2522–2531.

Mermillod, M., Vermeulen, N., Lundqvist, D., & Niedenthal, P.M. (2009). Neural computation as a tool to differentiate perceptual from emotional processes: The case of anger superiority effect. *Cognition*, *110*, 346–357.

Meschyan, G., & Hernandez, A. E. (2002). Is native language decoding skill related to second language learning? *Journal of Educational Psychology*, *94*, 14–22.

Monaghan, P., & Ellis, A. W. (2010). Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language*, *63*, 506–525.

Moore, V. (2003). An alternative account for the effects of age of acquisition. In P. Bonin (Ed.), *Mental lexicon. ''Some words to talk about words''* (pp. 67–93). Hauppauge, NY: Nova Science Publisher.

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

Pallier, C., Dehaene, S., Poline, J.-B., LeBihan, D., Argenti, A.-M, Dupoux, E., & Mehler, J. (2003). Brain imaging of language plasticity in adopted adults: Can a second language replace the first? *Cerebral Cortex*, *13*, 155–161.

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

Raman, I. (2006). On the age-of-acquisition effects in word naming and orthographic transparency: Mapping specific or universal? *Visual Cognition*, *13* (7/8), 1044–1053.

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

Seidenberg, M. S., & Zevin, J. D. (2006). Connectionist models in developmental cognitive neuroscience: Critical periods and the paradox of success. In Y. Munakata and M. Johnson (Eds.), *Processes of change in brain and cognitive development. Attention and Performance XXI*. Oxford, UK: Oxford University Press.

Stewart, N., & Ellis, A. W. (2008). Order of acquisition in learning perceptual categories: A laboratory analogue of the age of acquisition effect? *Psychonomic Bulletin & Review*, *15*, 70–74.

Ventureyra, V., Pallier, C., & Yoo, H.-Y. (2004). The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics*, *17*, 79–91.

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

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.

Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory and Cognition*, *32*, 31–38.

Zevin, J. D., Seidenberg, M. S., & Bottjer, S. W. (2004). Limits on reacquisition of song in adult zebra finches exposed to white noise. *Journal of Neuroscience*, *24*, 5849–5862.