A computational model of reading across development: Effects of literacy onset on language processing

Ya-Ning Chang[a,e,*], Padraic Monaghan[a,b], Stephen Welbourne[c,d]

[a] Department of Psychology, Lancaster University, UK
[b] Max Planck Institute for Psycholinguistics, the Netherlands
[c] Neuroscience and Aphasia Research Unit, University of Manchester, UK
[d] Sparx, Oxygen House, Exeter, UK
[e] MRC Cognition and Brain Sciences Unit, University of Cambridge, UK

ABSTRACT

Cognitive development is shaped by interactions between cognitive architecture and environmental experiences of the growing brain. We examined the extent to which this interaction during development could be observed in language processing. We focused on age of acquisition (AoA) effects in reading, where early-learned words tend to be processed more quickly and accurately relative to later-learned words. We implemented a computational model including representations of print, sound and meaning of words, with training based on children's gradual exposure to language. The model produced AoA effects in reading and lexical decision, replicating the larger effects of AoA when semantic representations are involved. Further, the model predicted that AoA would relate to differing use of the reading system, with words acquired before versus after literacy onset with distinctive accessing of meaning and sound representations. An analysis of behaviour from the English Lexicon project was consistent with the predictions: Words acquired before literacy are more likely to access meaning via sound, showing a suppressed AoA effect, whereas words acquired after literacy rely more on direct print to meaning mappings, showing an exaggerated AoA effect. The reading system reveals vestigial traces of acquisition reflected in differing use of word representations during reading.

Effects of learning on cognitive processing

Models of cognitive processing tend to assume that behaviour is based on responding to the current environment of the participant. For instance, behavioural analyses of processes involved in reading have focused on predicting responses based on properties of word stimuli that are observable in the current environment of the reader (Adelman & Brown, 2008; Adelman, Sabatos-DeVito, Marquis, & Estes, 2014; Balota et al., 2004, 2007; Brysbaert, Lange, & van Wijnendaele, 2000; Cortese & Khanna, 2007; Cortese & Schock, 2013; Yap, Balota, Sibley, & Ratcliff, 2012). Computational models of reading have thus focused on, and effectively addressed, these effects of the current language environment on reading (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Harm & Seidenberg, 1999, 2004; Perry, Ziegler, & Zorzi, 2007, 2010; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Spieler & Balota, 1997).

However, the assumption that processing is based on the configuration of stimuli in the immediate environment omits the potential effects that the personal history of the reader may have on their processing of current stimuli. Whereas this may be a useful simplifying assumption, making assessment of the reader's exposure tractable as a predictor of their reading performance, this simplification may also mask important effects of the learner's cognitive development that can help to explain current behaviour.

We know that early experience exerts a profound effect on the individual in terms of social, economic, and cognitive development (Belsky, Barnes, & Melhuish, 2007), and this impact of early experience has led to large-scale social programmes for early intervention to mitigate against negative effects on long-term life outcomes (e.g., USA: The White House, 2013; UK: Harold, Acquah, Sellers, & Chowdry, 2016). Such social programmes are based on evidence that social, economic, and academic performance measures in early childhood have substantial effects on quantifiable life outcome measures in later life (Bates et al., 1988; Hart & Risley, 1995; Rowe & Goldin-Meadow, 2009). Yet, the precise effects on representations and processing within the individual are not discernible from these studies: it is known that

early experience affects later performance, but less is known of how the cognitive system is shaped as a consequence of this early experience. Nevertheless, these effects of early experience provide substantial evidence that an adequate model of processing must take into account the life experience of the learner in explaining behaviour.

Cognitive processing is affected not only by the chronological properties of a potentially changing environment that the learner has experienced, but also by the changing architecture of the cognitive system that is required to respond to these environmental changes (Elman, 1990, 1993; Mareschal et al., 2007). In the cognitive sciences, two mechanisms have been proposed to account for the processes underlying the impact of early experience. First, it is proposed that later experiences are constructed on the back of early representations, such that later representations are influenced by earlier stored information (Anderson & Cottrell, 2001; McCloskey & Cohen, 1989; Smith, Cottrell, & Anderson, 2001). Second, it is proposed that in early development there is greater plasticity of the neural substrate that stores and processes information, meaning that early exposure results in greater dedication of resources to encode the experience (Ellis & Lambon Ralph, 2000; Westermann et al., 2007). Multidisciplinary methods have converged to provide a rich view of early experience affecting processing (Thomas & Knowland, 2009), but these studies have principally focused on very early development (Richardson & Thomas, 2008; Thomas & Johnson, 2008). A full understanding of the effect of early experience on life outcomes requires a perspective over the lifespan, revealing how early experience can continue to influence later processing.

Steyvers and Tenenbaum (2005) model of semantic associates elegantly demonstrated how early experience can have a profound effect on representation in the language domain. They observed that the age at which a word is acquired – its "Age of Acquisition" (AoA) – related to the number of semantic associates that participants produced in a free association task. In this task, participants are required to produce all words that are associated with a given target word. Words with earlier AoA tended to have more semantic associates than words with later AoA. Steyvers and Tenenbaum (2005) interpreted this in a small-scale model where semantic associative memory was gradually constructed, such that associations between concepts were built upon the network of concepts that were already in place. Resulting from this, those words that were first acquired tended to have more words associated with them than words that entered into the network at a later point. AoA effects in semantic associative memory were thus characterised as architectural – as a consequence of the chronological construction of the model.

Age of acquisition effects

Developmental models have provided valuable theoretical advances in determining how development and experience interact in forming the cognitive system (Mareschal et al., 2007; Thomas & Johnson, 2008). However, longer-term effects of development in affecting behavioural performance in cognitive tasks have been less studied. What is required to unfold the effect of early experience on later processing is a domain where the temporal nature of input has a well-documented and precise characterisation in terms of effects on behaviour. One such example that has been extensively studied is observations of age of acquisition (AoA) effects in lexical processing.

Some words are learned before others in children's language development. Which word is acquired first is partly due to frequency – words that tend to be acquired earlier tend to be more frequent in children's language experience – and though AoA is related to other psycholinguistic variables, such as frequency, semantic or orthographic neighbourhood size, word length, and concreteness – it cannot be reduced entirely to these other psycholinguistic variables (Brysbaert, 2017). It thus remains an important topic to determine how early experience affects language processing. There has been substantial progress in determining the impact of early experience on the domain of language processing, making this topic particularly suitable for investigation because of the abundance of observational data available (Alario et al., 2004; Barry, Johnston, & Wood, 2006; Barry, Morrison, & Ellis, 1997; Bates, Burani, D'Amico, & Barca, 2001; Bonin, Barry, Méot, & Chalard, 2004; Bonin, Chalard, Méot, & Fayol, 2002; Bradley, Davies, Parris, Su, & Weekes, 2006; Brown & Watson, 1987; Brysbaert, van Wijnendaele, & Deyne, 2000; Butler & Hains, 1979; Carroll & White, 1973; Chalard, Bonin, Méot, Boyer, & Fayol, 2003; Coltheart, Laxon, & Keating, 1988; Cortese & Khanna, 2007; Cuetos, Ellis, & Alvarez, 1999; Davies, Wilson, Cuetos, & Burani, 2014; Ellis & Morrison, 1998; Gerhand & Barry, 1999; Ghyselinck, Custers, & Brysbaert, 2004; Ghyselinck, Lewis, & Brysbaert, 2004; Gilhooly & Logie, 1980; Hernandez & Li, 2007; Hodgson & Ellis, 1998; Izura & Ellis, 2002, 2004; Johnston & Barry, 2006; Juhasz & Rayner, 2003; Lake & Cottrell, 2005; Lewis, Gerhand, & Ellis, 2001; Meschyan & Hernandez, 2002; Morrison & Ellis, 1995, 1999; Morrison, Hirsh, Chappell, & Ellis, 2002; Stadthagen-Gonzalez & Davis, 2006; van Loon-Vervoorn, van Ham, & van der Koppen, 1988; Wilson, Ellis, & Burani, 2012). In language processing, the effect of acquisition of a word provides a key empirical demonstration of the lifespan effects of the cognitive system's personal history of learning over the lifespan. The AoA of a word affects the individual's speed and accuracy at accessing, reading, and remembering words (Juhasz, 2005), such that early-acquired words are accessed more quickly and represented with greater fidelity than later-acquired words. The AoA effect therefore provides a unique window into the effect of a life history of vocabulary acquisition. It is not only a predictor of the accessibility of words in everyday language use, but also predicts their endurance when a speaker's language is depleted, as a consequence of ageing (Hodgson & Ellis, 1998) or in acquired dysfunctions such as aphasia (Bradley et al., 2006) or Alzheimer's disease (Holmes, Jane Fitch, & Ellis, 2006).

Explanations for AoA effects in lexical processing

There are two theoretical explanations for the AoA effect in lexical processing, aligning with the two broad theories for how early experience impacts cognitive development. First, AoA effects have been claimed to be a consequence of the incremental construction of semantic representations, whereby later acquired words are incorporated into a representation already containing early acquired words (Brysbaert & Ghyselinck, 2006; Li, Farkas, & MacWhinney, 2004; Steyvers, Shiffrin, & Nelson, 2004). From this perspective, early AoA words have a processing priority because they have richer, more embedded semantic representations than later AoA words (Steyvers & Tenenbaum, 2005). The alternative theoretical explanation for the AoA effect is that it is due to early plasticity on the learning of mappings between written, spoken, and semantic forms of the vocabulary (Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006; Marchman, 1993). Under this explanation, an AoA effect is expected particularly when the mappings between inputs and outputs are arbitrary, because they require greater computational resources to resolve the mapping (Zevin & Seidenberg, 2002).

To learn to recognise spoken words, one must acquire mappings between phonology and semantics of words, and to learn to read, a mapping between orthography and phonology and semantics must be constructed. For learning early acquired words, there is plasticity in the system, because the mappings are unconstrained by previous learning. However, for learning words later, the mappings must be accommodated around pre-existing mappings that have already been acquired. Computational models based on associative learning mechanisms have been successful in demonstrating that AoA effects can be observed as a consequence of changing plasticity in the learning system, which in turn relates to the fidelity of the mapping between representations for words (Ellis & Lambon Ralph, 2000; Monaghan & Ellis, 2010; Zevin & Seidenberg, 2002, 2004). Furthermore,

computational models have also demonstrated the varying effects of AoA depending on the characteristics of the mapping between orthography and phonology and orthography and semantics (Lambon Ralph & Ehsan, 2006). However, these previous models have not simultaneously taken into account phonological, orthographic and semantic representations engaged in reading, and the dynamics of a highly interactive system are not predictable from simulations of single mappings within models.

An emerging view seems to be that both the representation (prior learning influences processing) and the mapping (early plasticity influences processing) theory contribute to AoA effects (Menenti & Burani, 2007). For instance, lexical decision tasks demonstrate enhanced AoA effects compared to word naming tasks (Brysbaert & Ghyselinck, 2006), and in their review, the AoA effects are interpreted in terms of greater involvement of semantic representations. Yet, AoA effects can also be observed when semantics are not directly implicated, for tasks such as word naming (Monaghan & Ellis, 2002), consistent with the mapping theory, but only consistent with the representation theory if semantics is also involved indirectly in word naming. A more recent review (Brysbaert & Ellis, 2016) confirmed the larger effect of AoA for lexical decision than word naming, but also highlighted that a similar larger effect of frequency is also found for lexical decision than word naming. This suggested that AoA and frequency effects are in tandem to a certain degree and may have a common origin either in mappings or in representations. There are intriguing possibilities that the involvement of semantics in reading may vary during vocabulary development, both in terms of the relative ease of learning mappings between orthography and phonology compared to orthography and semantics, but also because over time the semantic representations become enriched (Li et al., 2004).

Our aim in this paper is to show the experience of learning is a vital part of the explanation, and not only the description, of cognitive processing, taking reading as an illustrative example of the individual's life history of experience reflected in their cognitive processing.

Modelling reading development

There are different architectural approaches available to develop computational models of reading. Dual-route traditions of modelling have been effective in simulating detailed data in word and nonword naming reaction times and accuracy of responses, such as the DRC (Coltheart et al., 2001) and the CDP+ (Perry, Ziegler, & Zorzi, 2007). However, these models are not yet able to account for the gradual development of both lexical and sublexical processing of word naming, and they have not yet implemented large-scale semantic representations interacting with orthographic and phonological representations (Nation, 2009; Taylor, Duff, Woollams, Monaghan, & Ricketts, 2015). Alternatively, computational models in the triangle modelling tradition (Harm & Seidenberg, 2004; Plaut et al., 1996; Seidenberg & McClelland, 1989) have demonstrated that the general framework of parallel distributed processing could be used to investigate different types of learning such as the effect of continuous (but non-incremental) experience in first and second language learning (Monaghan, Chang, Welbourne, & Brysbaert, 2017) and learning to read (Chang & Monaghan, 2019; Monaghan & Ellis, 2010). However, these models have not yet been applied to investigate the effect of lifetime experience on representation and processing in the reading system that incorporates interactivity among the three key representations involved in reading of orthography, phonology and semantics. Similar to the procedures used in Monaghan and Ellis (2010), the model we constructed learns gradually and incrementally from its exposure to words. Nevertheless, with the inclusion of the semantic system in the model, the learning involves more than effects along a single route. Specifically, the model initially learns to map between phonological and semantic representations of a subset of words in the vocabulary, to reflect pre-literate language exposure. Then, the model is exposed to written

words that are presented according to age-appropriate frequency between the ages of 5 and 18, and is required to learn to produce the phonological and semantic representations for each written word. This gradual, sequential training of the model implements the history of the learner in terms of their unfolding experience of the language, and opens up investigations of the learning of reading via direct and indirect pathways between representations, and how these might alter over the lifetime. We show that this interactivity of multiple routes and representations in the model has a fundamental importance in determining the locus of age of acquisition effects in reading, and further reveals complex effects in terms of changing division of labour along these pathways as reading develops.

There has been recent innovation in the extent to which models can be fit to behavioural data in order to account for variance in large-scale databases of word naming (e.g., Adelman & Brown, 2008; Adelman et al., 2014; Perry et al., 2007). Alternatively, a parallel tradition of modelling is to reflect the general principles, rather than to simulate precise fit, for human performance (Coltheart et al., 2001; Harm & Seidenberg, 2004; Monaghan & Ellis, 2010; Seidenberg & Plaut, 1998; Sibley, Kello, & Seidenberg, 2009). Despite these differences in aims, each approach to modelling reading must be able to reflect key behavioural phenomena in word reading. First, models must be able to learn the set of words in their environment to a high degree of accuracy, and also be able to generalise to accurately pronounce nonwords (Glushko, 1979; McCann & Besner, 1987; Whaley, 1978), to which the model has not previously been exposed (Coltheart et al., 2001; Harm & Seidenberg, 1999). Second, models should be able to reproduce key psycholinguistic effects of words, such as frequency (Brown & Watson, 1987; Brysbaert et al., 2000; Morrison & Ellis, 1995), neighbourhood size (Andrews, 1992), consistency of grapheme to phoneme correspondences (Hino & Lupker, 2000; Paap & Noel, 1991), and frequency by consistency interactions (Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987; Waters & Seidenberg, 1985).

These effects on word naming have previously been simulated in computational models of reading by observing the time-course and accuracy of representations in the region of the models corresponding to the phonology of the word (Chang, Furber, & Welbourne, 2012; Harm & Seidenberg, 1999; Plaut et al., 1996; Seidenberg & McClelland, 1989). Similar effects of psycholinguistic variables are also observed in lexical decision, but there is less convergence about how to simulate this task in models of reading, where it is sometimes taken as a discrimination task over phonological representations in a model (Pagliuca & Monaghan, 2006), or as a (proposed) activation of orthographic representations (Seidenberg & McClelland, 1989), or alternatively as a measure of polarity across either semantic representations (Plaut, 1997) or all the representations (Chang, Lambon Ralph, Furber, & Welbourne, 2013) associated with the words. Behaviourally, it is evident that a larger role of semantic representations is involved in generating such lexical decision responses, with a greater contribution of semantically-related measures of imageability and concreteness in predicting response times in lexical decision than in word naming (Cortese & Khanna, 2007; Cortese & Schock, 2013). Consequently, in analysing the model presented in this paper, we assessed the model's generation of semantic representations as a proxy for lexical decision responses. In addition to the psycholinguistic measures of frequency, neighbourhood size, and consistency (though a smaller, but significant, effect in lexical decision than word naming) we also determined whether the model could reflect concreteness, or imageability, effects observed in behavioural studies of lexical decision, in order to unpack the relative contributions of semantic variables and AoA effects in the model's performance (Coltheart et al., 1988; Cortese & Schock, 2013; Gilhooly & Logie, 1980; Strain, Patterson, & Seidenberg, 1995, 2002; van Loon-Vervoorn et al., 1988).

With the effects of learning history potentially discernible in the model's processing, various effects of AoA can also be explored in the model. Thus, earlier acquired words are quicker and more accurate to

name than later acquired words (Brown & Watson, 1987; Juhasz, 2005), and there should be a larger effect of AoA in lexical decision than in word naming (Brysbaert & Ghyselinck, 2006; Cortese & Khanna, 2007; Lambon Ralph & Ehsan, 2006). There is also an interaction between AoA and consistency observable in word naming (Monaghan & Ellis, 2002; Wilson et al., 2012). These effects were subsequently tested in the computational model. Simulating this range of effects in both word naming and lexical decision tasks provides an advance on previous modelling approaches to word processing in terms of breadth of coverage of behaviour.

AoA as a lense into the reading architecture

In the triangle model, a word is represented by an orthographic, phonological, and semantic representation, and activation passing between these representations. Prior to learning to read, the model can be trained on oral language skills, to simulate pre-literacy language exposure, acquiring mappings between phonological and semantic representations to simulate learning to speak and comprehend language (Harm & Seidenberg, 2004). The model then learns to map written forms of words onto these phonological and semantic representations. This can be done through learning to map directly along connections from orthography to phonology and orthography to semantics, but also through indirect mappings from orthography to semantics via phonology, or from orthography to phonology via semantics (see Fig. 1), reflecting theories of the reading architecture utilising dorsal (phonology-based) and ventral (semantics-based) language processing pathways (Ueno & Lambon Ralph, 2013).

For learning to read a known word, learning orthography to phonology is a quasi-regular mapping, which can be acquired with fewer computational resources than the arbitrary orthography to semantics mapping. Thus, it seems likely that reading a known word for meaning is more likely to engage the indirect orthography to semantics via phonology mapping, due to the availability of the phonology to semantics mapping that is previously acquired during oral language development. In contrast, an unknown written word's meaning must be learned either by acquiring the mapping directly from orthography to semantics, or by acquiring an orthography to phonology mapping, and a new phonology to semantics mapping. For words acquired post-literacy, then, the model is predicted to be more likely to rely on the direct pathway from orthography to semantics.

AoA effects have been shown to be much larger for tasks involving semantics (e.g., lexical decision), than for tasks involving production of phonology (Brysbaert & Ghyselinck, 2006; Cortese & Khanna, 2007;
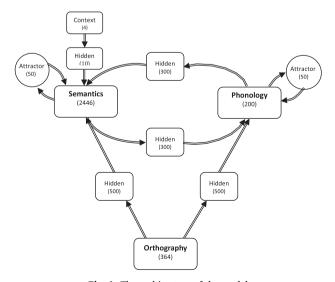
Lambon Ralph & Ehsan, 2006). Consequently, the size of the AoA effect can be used to indicate the extent to which the direct and indirect pathways from orthography to semantics are involved in reading. If the AoA effect is large, semantics is likely to be involved via the direct pathway from orthography to semantics, as a reflection of AoA effects emphasised in an arbitrary mapping (Lambon Ralph & Ehsan, 2006). If the AoA effect is small, then the direct mapping to semantics is likely to be less heavily involved, and the quasi-regular orthography to phonology mappings will be more prominent, which reduce the influence of AoA (Ellis & Lambon Ralph, 2000; Zevin & Seidenberg, 2002). Though there will still be an effect of AoA on the pre-literature phonology to semantics pathway, extensive training along this route is likely to have reduced its impact on behaviour. The present model can thus be applied to explore the effects of literacy on the architecture of the reading system in terms of pathways employed between pre-literacy and post-literacy acquired words.

In the next section, we present a developmental model of reading that learns dynamically as the reading environment unfolds. In Simulation 1, the model was trained with objective AoA measures based on the educator's word frequency guide (WFG) by Zeno, Ivens, Hillard, and Duvvuri (1995) to mimic incremental learning of natural reading development. We established whether the model was able to reproduce AoA effects in word naming and lexical decision, in addition to other standard effects (such as frequency and consistency), to investigate whether the model's processing was sensitive to its history of learning. To demonstrate that the effect of AoA was a consequence of incremental learning and not confounded properties of AoA, Simulation 2 trained a model identical to that in Simulation 1 except that we randomised each word's AoA in such a way that AoA was completely independent of other lexical variables. We predicted that the patterns of AoA in word naming and lexical decision would be similar to those in Simulation 1.

We then tested how the various pathways in the model – either direct from orthography to phonology or semantics, or mediated via phonology-semantics processing – operated during reading development, and in particular whether there were distinct patterns of processing for words to which the model had been exposed prior to onset of literacy and those for which phonological, semantic, and orthographic representations were all first experienced by the model post-literacy. Together these analyses provide insight into the generation and change of the reading system in particular, and cognitive processing more generally, as a consequence of experience. In the final section of the paper, we test the model's predictions of vestigial effects of literacy onset on adult reading processing.

Simulation 1: A developmental model of reading

Network architecture

The architecture of the model is shown in Fig. 1. The model was based on the triangle model of reading by Harm and Seidenberg (2004). The model consisted of three key processing layers including orthographic, phonological and semantic layers, and five hidden layers for intermediation between the processing layers.

An attractor layer, which contained 50 units, was connected to and from the phonological layers. Similarly, there was a set of 50 attractor units for the semantic layer. The use of attractors was to help the model to develop stable phonological and semantic representations of words. In addition, there were four context units connecting to the semantic layer through a set of 10 units. The context units provided additional information when presenting the model with homophones. One context unit was active for each homophone. But for words within the same homophone family, different context units were randomly assigned. In this way, each context unit was almost equally active across the training corpus. For non homophones, none of the context units was active. Pilot simulations demonstrated that the intermediary connections between context units and the semantic units were required for accurate

Fig. 1. The architecture of the model.

learning, suggesting some non-linearity in the semantic representations for the set of homophones.

The semantic layer was connected to the phonological layer through a set of 300 hidden units, and the phonological layer was connected back to the semantic layer through another set of 300 hidden units. The orthographic layer was connected to both the phonological and semantic layers through different sets of 500 hidden units.

Representations

The schemes of orthographic, phonological and semantic representations were the same as those used in Harm and Seidenberg (2004) model. The training corpus contained 6229 monosyllabic words, which covered most monosyllabic words and their inflected forms used in English. Frequency of each word was derived from the Wall Street Journal corpus (Marcus, Marcinkiewicz, & Santorini, 1993), and the score was log-transformed.

For orthography, each word was represented by 14 letter slots, and each slot comprised 26 units, one for each of the 26 letters of the alphabet. Words were positioned with their first vowel aligned on the fifth slot. For words having two adjacent vowels, the second vowel was placed on the sixth slot; otherwise, all the units in that slot were not active. Consonants preceding or following the vowel(s) were positioned in adjacent slots to the vowel(s) (so, for example, *yes* was represented as _ _ _ y e _ s _ _ _ _ _ _ _, and *great* as _ _ g r e a t _ _ _ _ _ _).

For phonology, each word was represented by eight phoneme slots, with each slot consisting of 25 phonological features. Each word was positioned with its vowel at the fourth phoneme slot. The first three slots were for onset consonants, and the last four slots were for coda consonants (so *yes* was _ _ y E s _ _ _ and *great* was _ g r eI t _ _).

The semantic representation for each word was derived from Wordnet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990) following Harm and Seidenberg (2004). Each semantic representation was composed of 2446 semantic features. The presence of semantic features was encoded as one, and the absence of semantic features was encoded as zero. Representations derived from Wordnet have validity in reflecting judgments about semantic similarity of concepts (e.g., Maki, McKinley, & Thompson, 2004), but future research with semantic representations that more closely reflect behavioural measures of semantic similarity, such as semantic priming effects, could further improve the ability of the model to reflect semantic relations (Mandera, Keuleers, & Brysbaert, 2017). The Wordnet representations that we used reflect a mature semantic system, and alternative approaches can add further insight into language acquisition by implementing the development of semantic representations (e.g., Li et al., 2004). However, note that the semantic representations are employed as a target, and the model begins by learning approximations to the fully-specified semantic system, acquiring more nuanced distinctions between semantic representations as training proceeds. In this respect, the model gradually acquires a semantic system.

Training and testing

The training process had two phases. In oral language training, the model was trained with the mappings between phonology and semantics. This phase of training was an attempt to mimic the fact that children generally have developed some language skills (e.g., speaking and hearing) before learning to read. In the reading development phase, the full reading model was trained. For both training phases, the training parameters were kept the same. The model was trained with a learning rate of 0.05 using a back-propagation through time (BPTT) algorithm with input integration and a time constant of 0.33 (Harm & Seidenberg, 2004; Plaut et al., 1996). The weight connections were updated on the basis of cross-entropy errors computed between the target and the actual activation of the output units.

In oral language training, the model was trained on a speaking task,

learning mappings from semantic to phonological representations, and a hearing task, learning mappings from phonological to semantic representations. The model also learned to develop a stable phonological attractor, learning mappings from phonological to phonological representations, and a stable semantic attractor, learning mappings from semantic to semantic representations. For both the speaking and hearing tasks, the input pattern of each word was clamped and presented for eight time samples, and in the last two time samples, the model was required to reproduce the target pattern of the word. For both the phonological and semantic attractors, the input pattern of each word was clamped and presented from the first time sample, and then in the last two time steps, the model had to reproduce the target pattern of the word. As semantic representations for each word were distinctive from each other, the model could learn the speaking task without additional information. So the input of context units was supplied only for the hearing task. During training, the four tasks were interleaved with 40% of trials for the speaking task, 40% of trials for the hearing task, 10% of trials for the phonological attractor and the remaining 10% for the semantic attractor.

During oral language training, the model was exposed to 2738 monosyllabic words, which were the most common words occurring in reading materials before age 18 based on the educator's word frequency guide (WFG) by Zeno et al. (1995). The WFG describes frequencies of words from age five to adult, and so the measure of AoA in the model begins at age five when literacy training begins for the model. In order to isolate the effects of AoA in literacy experience, we presented the model with a subset of this set of words, comprising approximately half of the entire vocabulary, during the pre-literacy training. During this oral language training, words were sampled according to their frequency (but were not staged in their presentation according to age), and this meant that not all the words were acquired simultaneously, or acquired accurately, prior to literacy training commencing. Note that children vary widely in the quality and diversity of vocabulary they hear (e.g., Rowe, 2012), and so it is not straightforward to determine a representative pre-literacy exposure for children. This allowed the model to be exposed to a certain range of words including some very high and low frequency words, though only the higher frequency words would be learned accurately, and it meant that the AoA effects in literacy could be related to the model's experience of staged reading materials.

In the reading development phase, the model was trained on the reading task, which was to learn the mappings from orthography to both semantics and phonology, along with the four tasks in the oral language training phase. Following Monaghan and Ellis (2010), the model was trained with a cumulative process of learning to read, to reflect 14 reading stages, one for each year based on by Zeno et al.'s WFG (1995). The words in WFG were graded into 13 different grade levels by using readability measures, corresponding to the age range from five to 18 in the American and British schooling systems, and the words appeared in adulthood were presented at stage 14. The points for words entered into training were dependent on their first occurrence in the WFG with a frequency greater than a certain threshold. The threshold scheme was adopted from Monaghan and Ellis (2010), where the cut-off frequency reduced with each grade to ensure that the model learned incrementally, mimicking children's learning to read. As can be seen in Table 1, the model started to learn a small set of the most frequent words occurring in age-appropriate texts at stage one, and gradually more and more words were learned over the time course of learning. At stage 14, the entire vocabulary was presented to the model.

For the reading task, the orthographic representation of a word along with the context layer representation were clamped and presented for 12 time samples, and from time samples six to 12, the model was required to produce the phonological and semantic representations for that word. All the five tasks were interleaved during training, but the ratio for each task except the attractor tasks varied as the training proceeded. The training ratios for both the hearing task and speaking

Table 1
The training paradigm for both oral language training and reading training.

	Number of words for SP, PS, and Atts	Number of words for OP and OS	Training ratio (SP/PS/Atts/ (OP, OS))	Number of training presentations
Oral language				
	2738	0	0.4/0.4/0.2/0	600 K
Reading				
1	2738	103	0.4/0.4/0.1/0.1	100 K
2	2738	535	0.35/0.35/0.1/0.2	50 K
3	2738	775	0.3/0.3/0.1/0.3	60 K
4	2738	1107	0.25/0.25/0.1/0.4	50 K
5	2738	1675	0.2/0.2/0.1/0.5	40 K
6	2738	1833	0.2/0.2/0.1/0.5	80 K
7	2738	2183	0.2/0.2/0.1/0.5	80 K
8	2738	2322	0.2/0.2/0.1/0.5	80 K
9	2738	2492	0.2/0.2/0.1/0.5	80 K
10	2738	2695	0.2/0.2/0.1/0.5	80 K
11	2738	2996	0.2/0.2/0.1/0.5	80 K
12	2738	3307	0.2/0.2/0.1/0.5	80 K
13	2738	2738	0.2/0.2/0.1/0.5	80 K
14	6229	6229	0.2/0.2/0.1/0.5	400 K

Note. SP: semantics - > phonology; PS: phonology - > semantics; Atts: phonological and semantic attractors; OS: orthography - > semantics; OP: orthography - > phonology; K: 1000 training trials.

task gradually decreased from 40% to 20% in steps of 5% during literacy training, while the training ratio for the reading task gradually increased from 10% to 50% in steps of 10%. The details of the training regime can be seen in Table 1. The number of presentations in each reading training stage was adopted from those used in Monaghan and Ellis (2010) but was adjusted to accommodate the gradually increasing reading training ratios in the present study to ensure accurate learning. All the other training procedures remained the same as in oral language training. After the cumulative presentation of words according to age, the model was trained for 400,000 presentations with the entire vocabulary to simulate adult reading. At this point, the performance was close to converging on accurate phonological and semantic productions

of written words, and we examined the model's performance at this stage to simulate young adult reading performance.

After oral language training, the model was tested on both the speaking and hearing tasks. For the speaking task, the semantic representation of a word was presented, and the activation of units at the phonological layer at the end of the eight time samples was recorded. Error score was measured as the sum of the squared differences between the activation of each input unit and its target activation. The accuracy of the model's phonological production was determined by whether, for each phoneme slot, the closest phoneme to the model's actual production was the same as its target phoneme. For the hearing task, the phonological representation of a word was presented, and the activation of units at the semantic layer at the end of the eight time samples was recorded. Error score was measured as the sum of the squared differences over the semantic layer. The semantic accuracy was measured by computing the Euclidean distance between the model's actual semantic representation and the semantic representation of each word in the training corpus. If the smallest distance was for the target representation, then the model was judged to be correct.

After reading training, the model's reading performance was tested. The orthographic representation of a word was presented, and the activation of units at both the semantic layer and the phonological layer at the end of the 12 time samples were recorded. The measurement of error score and accuracy for both semantic and phonological outputs were the same as in the oral language training phase.

Results

We first established whether the model was able to reproduce key behavioural phenomena associated with single word reading and AoA effects (Monaghan & Ellis, 2010).

Word reading accuracy

Oral language training was halted after 0.6 million epochs where the model achieved an accuracy rate of 97.85% on the speaking task and an accuracy rate of 93.35% on the hearing task. Fig. 2 shows the performance of the reading model over the time course of the reading training. The accuracy rates indicate the proportions of all words
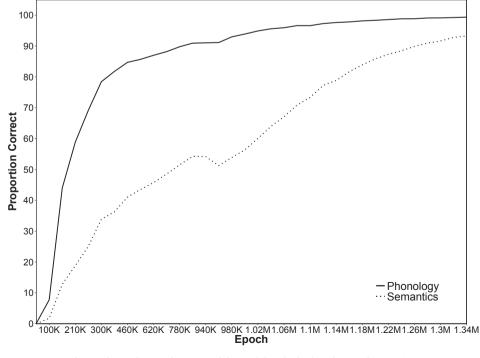

Fig. 2. The reading performance of the model on both phonology and semantics.

generated correctly at each time epoch for both phonology and semantics. The resulting patterns showed learning curves for English reading similar to previous computational studies of reading (e.g., Harm & Seidenberg, 2004) where phonology was learned faster than semantics. At the end of the 14th stage of training (i.e., 1.34 million presentations), the performance of the model on both phonology and semantics gradually reached an asymptote, in which the model accurately produced 99.36% of phonological representations and 93.27% of semantic representations on the reading task.

Nonword reading accuracy

To evaluate if the model could generalise to read previously unseen words, we tested the performance of the model on nonword reading. The nonword set included 86 pseudowords from Glushko (1979), where half of them were derived from words that had consistent spelling-to-sound mappings, and the other half were derived from words that had inconsistent spelling-to-sound mappings. The nonword set also included 80 pseudowords taken from McCann and Besner (1987). The model generalised well to nonwords and was able to pronounce 80.72% (SD = 1.49) of nonwords correctly. This result is broadly comparable with that of Harm and Seidenberg (2004), whose model correctly pronounced 86.67% of the same set of nonwords. The resulting performance is also close to that of Monaghan and Ellis (2010) simulation of incremental reading training in a model with only orthography and phonology implemented, which correctly pronounced 84.6% (SD = 2.5) of the same set of nonwords but with an additional 80 homophones from McCann and Besner (1987) also included.

The model was thus able to learn to read words accurately and generalise to pronounce novel words effectively.

AoA and psycholinguistic variable effects

We next examined whether the model was able to produce AoA effects in both word naming and lexical decision along with other key psycholinguistic factors, including frequency and consistency, affecting reading performance.

In the model, AoA effects may reside either in the semantic representations, or in the mappings between representations (Brysbaert & Ghyselinck, 2006), or both. According to the semantic representation of AoA, early-learned words tend to have richer semantic representations compared to late-learned words (Brysbaert & Ghyselinck, 2006). Thus, AoA should be related to the complexity of semantic representations. To measure this possibility, we investigated to what degree AoA could account for semantic richness in the representations derived from Wordnet (Miller et al., 1990). We assumed that semantic richness was reflected in the number of semantic features (NoF) in the representations (Plaut & Shallice, 1993). In addition, Grondin, Lupker, and McRae (2009) demonstrated that *types* of semantic features mattered in addition to *number* of features, where shared features among words could be the key to explain semantic richness effects, and so we derived a further measure of semantic feature properties from the semantic representations: semantic neighbours (SemN), which measured the degree of overlap among semantic representations. This latter variable was computed by determining the cosine similarities between each word and all other words, and then taking the average score of the top five highest cosine values. Note that the decision to use the top five cosine values is arbitrary but the key was to quantify how the semantic features for a word are shared with those of other closely-related words. A high score for a word meant that its neighbours share a similar set of semantic features, thus occupying a dense, overlapping region of the semantic space.

Correlation analyses were conducted with both SemN and NoF as well as AoA (see Table 2). The results showed that AoA was significantly correlated with SemN, $r = 0.17$, $p < .001$ while the correlation between AoA and NoF was not significant. This finding suggested that AoA was related to the overlapping features within semantic representations in the model.

Word naming and lexical decision tasks were simulated by mappings from orthographic to phonological representations (Chang et al., 2012; Monaghan & Ellis, 2010), and by mappings from orthographic to semantic representations (similar to the polarity measure used in Plaut, 1997), respectively. According to the mapping theory of AoA, we would expect to obtain a larger AoA effect in lexical decision (semantics) than in word naming (phonology), after semantic richness variables are included in the analysis. If this is found to be the case, then both semantic representations and mappings are driving the AoA effects in the model.

To compare the simulation results to the behavioural findings of psycholinguistic effects reported by, for example, Cortese and Khanna (2007), multiple regression analyses were conducted on mean phonological and semantic error scores produced by the model for each word to examine the AoA effects in the model, using a similar set of lexical variables to behavioural studies. Error scores for phonology and semantics were used to provide finer-grained measures than accuracy to reflect the ease with which the model produces the representation from the written word. Model error measures have been used to correspond to behavioural response times for word naming and lexical decision (e.g., Harm & Seidenberg, 1999; Monaghan et al., 2017; Monaghan & Ellis, 2010; Plaut et al., 1996). The predictors were cumulative frequency (CF), which was the actual frequency of occurrence of words during the model's training; orthographic neighbourhood size (OrthN), which was the number of words that can be made by changing one letter of the target word (Coltheart, Davelaar, Jonasson, & Besner, 1977); orthographic word length (Len); consistency (Cons), which was a measure of rime consistency, computed by determining the number of friends (sharing the same rime and pronunciation) divided by the total number of words sharing the same rime and weighted by their frequency values from the words in the model's training corpus; and concreteness (Conc), taken from Brysbaert, Warriner, and Kuperman (2014). AoA was taken as one of the 14 reading stages during the training of the model derived from the WFG (Zeno et al., 1995). We also included both number of features (NoF) and semantic neighbour (SemN), to consider both representational and mapping effects of AoA simultaneously. Note that some recent studies have shown that contextual diversity is a better predictor in accounting for lexical processing compared to frequency measures (Adelman, Brown, & Quesada, 2006; van Heuven, Mandera, Keuleers, & Brysbaert, 2014). It might be argued that AoA effects could diminish when contextual diversity is considered. However, Davies, Arnell, Birchenough, Grimmond, and Houlson (2017) recently investigated variation in psycholinguistic effects across the lifespan including both AoA and contextual diversity. Although contextual diversity in Davies et al. (2017) served as an estimate of frequency rather than as a measure of diversity in experience, they demonstrated a reliable AoA effect when contextual diversity was controlled for. Relatedly, Hsiao and Nation (2018) showed that a measure of semantic diversity (Hoffman, Lambon Ralph, & Rogers, 2013) proved a complementary predictor, rather than a replacement, for AoA in relation to children's reading performance.

The correlations between predictors and dependent variables (mean error scores (MES) by the model at the end of the 12 time samples for phonology and semantics) are shown in Table 2. As expected, AoA was negatively correlated with CF and Conc, suggesting early learned words tend to be high in frequency and imageability. OrthN was negatively correlated with Len, reflecting that longer words tend to have fewer neighbours. Table 2 also demonstrates CF was negatively correlated with both phonological MES and semantic MES produced by the model, indicating that the key effect of frequency in word naming and lexical decision were both realised in the model's performance, with higher frequency words resulting in more accurate responses in both phonology and semantics, simulating word naming and lexical decision, respectively. Similarly, AoA was positively correlated with both phonological MES and semantic MES, demonstrating an effect of AoA in the model: earlier experienced words resulted in more accurate responses in the model. As for Cons, it was positively corrected with phonological

Table 2
The correlations between predictors and the dependent variables.

		CF	Len	OrthN	Cons	AoA	Conc	NoF	SemN
IV	CF	1							
	Len	−.24***	1						
	OrthN	.14***	−.64***	1					
	Cons	−.17***	.04***	−.02	1				
	AoA	−.84***	.21***	−.17***	.13***	1			
	Conc	−.02	−.01	.09***	.04**	−.05***	1		
	NoF	−.08***	−.06***	.06***	−.01	.02	.31***	1	
	SemN	−.34***	.06***	.06***	.07***	.17***	.40***	.39***	1
DV	MES (Phonology)	−.32***	.17***	−.26***	−.19***	.34***	−.02	.04***	.06***
	MES (Semantics)	−.45***	−.02	0	.05***	.51***	−.07***	−.02	−.08***

MES: mean error score.
***$p < .001$, **$p < .01$, *$p < .05$.

MES, suggesting more consistent words resulted in quicker naming responses in the model. Although Cons was negatively correlated with semantics MES, the correlation was considerably smaller, and the effect is eliminated when other variables are taken into account (see below).

Prior to regression analyses, words that the model misread, and words without measures for all psycholinguistic variables were omitted. In addition, outliers (> 3 standard deviations from the mean) were not considered, leaving 5213 words for phonology analyses and 5235 words for semantics analyses. Both the phonological and semantic error scores were log transformed to reduce the skew of performance distribution, and all the predictor variables were centred in order to explore interaction terms. The distributions after log transformation for both phonology (M = −3.18, SD = 0.49) and semantics (M = 0.051, SD = 0.065) are shown in Fig. 3. Hierarchical regression analyses were then conducted on phonological MES and semantic MES separately. For all regression models, collinearity diagnostic analyses showed all variance inflation factors (VIFs) smaller than 4, confirming no problematic multicollinearity.

For the word naming task, in step 1 all psycholinguistic variables were entered into the regression model except AoA (see Table 3). All of the variables except SemN made significant contributions. When AoA was entered into the regression model in step 2, it was a significant predictor, but Conc became a non-significant predictor, suggesting Conc and AoA share some common variance.

Similar analyses were conducted for the lexical decision task. In step 1, CF, Len, Conc, SemN and NoF were significant predictors (see Table 3). Again, in step 2, AoA was a significant predictor, and Conc became a non-significant predictor.

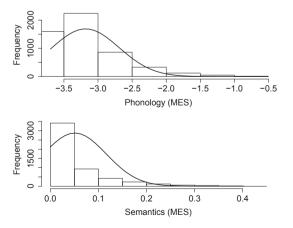These results showed that the AoA effects were found in both word

Fig. 3. The log distributions of mean error scores (MES) generated by the model for phonology (M = −3.18, SD = 0.49) and semantics (M = 0.051, SD = 0.065) with the normal distribution curves plotted based on their means and standard deviations.

Table 3
Results from two-step regression analyses for the exploration of AoA in predicting both word naming and lexical decision model performance.

		Word Naming	Lexical Decision
Step 1		Standardised β	Standardised β
	CF	−0.354***	−0.582***
	OrthN	−0.263***	0.024
	Cons	−0.247***	−0.013
	Len	−0.073***	−0.121***
	Conc	−0.052***	−0.075***
	NoF	0.042**	0.041**
	SemN	−0.013	−0.259***
	R^2 (%)	22.08%	28.82%
Step 2			
	AoA	0.192***	0.365***
		$\Delta R^2 = 0.96\%$	$\Delta R^2 = 3.46\%$

***$p < .001$, **$p < .01$, *$p < .05$.

naming and lexical decision even though the effect of semantic richness was also considered in the analysis. Also the standardised beta value (β) was larger for the lexical decision than for the word naming task, replicating behavioural studies showing a larger AoA effect in tasks involving semantics than phonology (Table 3). The results also confirm the observations of Brysbaert and Ellis (2016) review that both frequency and AoA demonstrate larger effects in lexical decision than word naming, suggesting that the sources of these effects are overlapping.

One point of divergence with the behavioural data, as shown in Table 3, was that the model produced lower error scores for longer words than shorter words in word naming when all other psycholinguistic variables had been taken into account. This was due to the correspondence between OrthN and Len in the current word set being highly related. Re-running the regression analyses without OrthN resulted in Len as a significant predictor in accord with behavioural data, $β = 0.095$, $t = 7.38$, $p < .001$, where longer words were processed less accurately than shorter words.

Further regression analyses were conducted to examine the interaction terms. Two interaction terms were created: CF × Cons, to determine whether the model can replicate the widely observed consistency by frequency interaction (Paap & Noel, 1991; Taraban & McClelland, 1987); AoA × Cons, to determine whether the model could reproduce the consistency by AoA effect (Monaghan & Ellis, 2002).

In step 1, all the variables including AoA were entered into the regression model, and in step 2, the interaction terms were entered into the model separately. The results are summarised in Table 4. For word naming, both CF × Cons (Fig. 4) and AoA × Cons (Fig. 5) were significant predictors, reproducing the key behavioural interaction effects with consistency. For lexical decision, both CF × Cons and AoA × Cons failed to reach significance. Thus, consistency effects were less

Table 4
Results from two-step regression analyses for three interaction terms in predicting both word naming and lexical decision model performance.

	Word Naming	Lexical Decision
Step 1: Lexical Variables, R^2	0.230	0.285
Step 2: Interactions		
Analysis 1:	0.052[***]	−0.003
CF × Cons, standardised β		
Analysis 2:	−0.067[***]	−0.010
AoA × Cons, standardised β		

[***] $p < .001$, [**] $p < .01$, [*] $p < .05$.

pronounced for lexical decision, consistent with the behavioural data (Andrews, 1982; Seidenberg et al., 1984; Waters & Seidenberg, 1985).

Multiple regression analyses on model performance demonstrated that AoA accounted for unique variance in both word naming and lexical decision, when other potentially confounding variables such as cumulative frequency and concreteness had been considered. In addition, the AoA effect was substantially larger in accounting for lexical decision than word naming responses. Collectively, then, the regression results are consistent with the findings of previous regression analyses for behavioural (Cortese & Khanna, 2007) and computational (Monaghan & Ellis, 2010) studies, and demonstrated that key behavioural effects of psycholinguistic factors influencing reading and lexical decision revealed in mega-studies were reproduced in the current model. These simulation results are consistent with theories of dual sources of AoA effects in the reading system: early-acquired words are prioritised in terms of both processing within the mappings and the richness of representations (Brysbaert & Ghyselinck, 2006).

Simulation 2: A developmental model of reading trained with randomised AoA

This simulation was similar to that of Simulation 1, except that AoA was randomised so that AoA was uncorrelated with other lexical semantic variables. We predicted similar AoA effects to those observed in Simulation 1, namely we would observe the influence of AoA on both

the model's word naming and lexical decision performance with a larger effect for lexical decision.

Network architecture and representations

The architecture and the representations were the same as in Simulation 1.

Training and testing

In this simulation, the AoA score was randomly assigned to each word in the training corpus. Training and testing were otherwise identical to Simulation 1.

Results

Fig. 6 shows the accuracy of the model for mapping from orthography to phonology and orthography to semantics over the time course of the reading training. The resulting patterns of learning were similar to those of Simulation 1, where phonology was easier to learn compared to semantics.

We next examined if the model trained with randomised AoA was able to produce a unique AoA influence on word naming and lexical decision when other psycholinguistic variables were controlled for. Table 5 shows the correlation between randomised AoA and other psycholinguistic variables. Randomised AoA was orthogonal to most of the variables, in particular word frequency. Although it was significantly correlated with SemN, the correlation was very small. However, there was a significant correlation between randomised AoA and cumulative frequency. This was because cumulative frequency measured the number of times that a word was presented to the model over the time course of training, which was unavoidably related to the first entry of the word into training.

As in Simulation 1, hierarchical regression analyses were conducted with phonological MES and semantic MES produced by the model for each word as dependent variables. In addition to the same set of predictors used in Simulation 1, we also included word frequency because
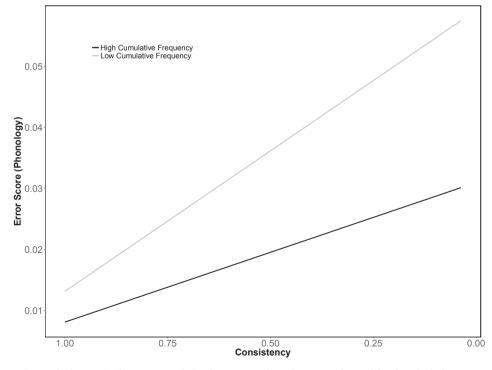

Fig. 4. The interaction between cumulative frequency and consistency on the model's phonological output.
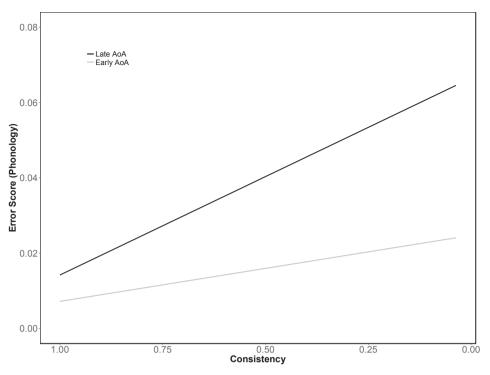
Fig. 5. The interaction between AoA and consistency on the model's phonological output.
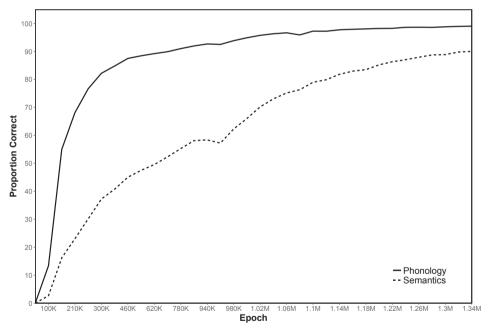

Fig. 6. The reading performance of the model trained with randomised AoA on both phonology and semantics.

Table 5
Correlations between randomised AoA and the other psycholinguistic variables.

	CF	WF	OrthN	Cons	Len	Conc	NoF	SemN
Randomised AoA	-0.54^{***}	0.00	0.01	0.00	0.01	0.03	0.01	0.04^{*}

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

it was designed to be uncorrelated with randomised AoA, and it would be critical to examine its influence. After incorrect items, outliers (> 3 standard deviations from the mean), and words without measures for all psycholinguistic variables were discarded, 5227 words for word

naming analysis and 5208 words for lexical decision were analysed. For word naming, in step 1 all predictors were entered into the regression model except AoA (see Table 6). The results showed that WF, CF, OrthN, Cons, Len, Conc, and SemN all made significant contributions. Importantly, when AoA was entered into the regression model in step 2, it was a significant predictor. But the sign of CF changed from negative to positive, indicating its shared common variance with AoA. For lexical decision, in step 1, all of the variables made significant contributions except that OrthN was marginally significant, and Cons was not significant (see Table 6). Again in step 2, AoA was a significant predictor while the sign of CF changed.

The resulting patterns of regression analyses were generally similar

Table 6
Results from two-block regression analyses for the exploration of AoA in predicting both word naming and lexical decision model performance.

		Word Naming	Lexical Decision
Step 1		Standardised β	Standardised β
	WF	−0.195***	−0.489***
	CF	−0.114***	−0.054***
	OrthN	−0.292***	0.030+
	Cons	−0.224***	0.006
	Len	−0.129***	−0.122***
	Conc	−0.083***	−0.139***
	NoF	0.006	0.055***
	SemN	0.039**	−0.134***
	R² (%)	19.25%	27.95%
Step 2			
	Randomised AoA	0.183***	0.213***
		Δ*R²* = 1.68 %	Δ*R²* = 2.27%

***p < .001, **p < .01, *p < .05, +p < .1.

to those in Simulation 1. Most importantly, when AoA in the model was orthogonal to word frequency and concreteness, which are correlated with AoA in natural language, there remained a reliable AoA effect on the model's word naming and lexical decision performance, demonstrating the importance of the sequence of word learning on the model's reading acquisition. The difference in the effect for word naming compared to lexical decision was smaller than that found in Simulation 1 with AoA derived from WFG, but the effect was in the same direction as before.

Effects of literacy onset on reading

Literacy is known to have profound effects on language processing, resulting in changes to phonological awareness (Hulme, Bowyer-Crane, Carroll, Duff, & Snowling, 2012; Morais, Cary, Alegria, & Bertelson, 1979), changes to phonological processing of words (Smith, Monaghan, & Huettig, 2014), as well as semantic fluency (Kosmidis, Tsapkini, Folia, Vlahou, & Kiosseoglou, 2004), and even visual processing (Szwed, Ventura, Querido, Cohen, & Dehaene, 2012). However, less studied are the potential effects of literacy on the architecture of the reading system in terms of pathways employed for different words according to how they are learned – whether from print or from oral language experience. Prior to literacy, the learner acquires mappings between sound and meaning representations of words, through listening and comprehending words, and speaking words for others' comprehension. However, once the child begins to learn to read for these already known words, mappings will be generated from print to the stored sound and meaning representations. But for new words, the print form will be mapped onto newly acquired sound and meaning representations, where the mappings between sound and meaning are not available in advance.

In terms of the operation of the reading system, this difference between pre-literacy and post-literacy acquired words is likely to be profound. In the triangle model of reading (Seidenberg & McClelland, 1989), there are two routes by which a printed word can be pronounced. This can occur directly, through learned mappings between print and sound, or indirectly from print via semantics to sound. Similarly, for word comprehension, the mapping from print can be directly to meaning, or indirectly, from print to sound to meaning. For words acquired prior to literacy, the indirect route is more likely to be available to support reading, because the sound to meaning routes are already acquired, whereas for words acquired post-literacy, the indirect route requires two mappings to be acquired.

Furthermore, the properties of the mappings from print to sound and meaning will also contribute to the extent to which direct and indirect mappings are utilised. Regular mappings, such as between print

and sound in English, are easier to acquire than arbitrary mappings, such as between print and meaning. Thus, for print to sound mappings, the direct route is more likely to be prioritised than the indirect route. By contrast, for print to meaning mappings, the indirect route is more likely to be prioritised than the direct route, at least for words acquired prior to literacy, where the sound to meaning mapping is already in place in the language processing system.

For words acquired prior to literacy, the indirect route is predicted to be more likely to have a greater influence on processing meaning, while for words acquired post-literacy onset, the direct route is likely to have a greater influence. In contrast, for print to sound mappings, as in word naming, no or only a small difference between pre- and post-literacy processing would be expected. This is because both pre- and post-literacy words will be mapped via fast-acquired direct print to sound mappings, though it is possible that the indirect route might have some influence on pre-literacy words.

The present model of Simulation 1, with the implementation of incremental learning, presents an ideal context for a computational test of the extent to which different processing routes operate for reading words acquired orally pre- versus post-literacy onset. In particular, the size of the AoA effect could be used to gauge the extent to which the reading pathways involving semantic access are utilised. Consequently, for word naming, a small difference between pre- and post-literacy processing in the AoA effect is expected. Whereas for lexical decision, a larger AoA effect for post-literacy words would be expected because of greater use of direct mappings from orthography to semantics compared to pre-literacy words that are more likely to be processed through an indirect route from orthography to semantics via phonology. It is, of course, the case that mappings between phonology and semantics are also arbitrary, but these mappings would exert a smaller AoA effect than that observed for the newly acquired mappings because they are intensively trained, and acquired earlier in acquisition, thus reducing distinctions between words due to greater plasticity of resources for early-learned mappings (Ellis & Lambon Ralph, 2000).

Words are acquired gradually prior to literacy onset (Brysbaert, 2017). However, in our modelling we did not introduce graded incremental learning of words prior to literacy. This was a design choice to ensure that observable AoA effects were related to the point at which the word was first read. Furthermore, use of the WFG measure to reflect AoA in reading acquisition is unsuited to determining AoA prior to literacy onset, because it reports exposure to words in written materials, from reading age one to 14 (corresponding to age five to 18). As our simulations were intended to isolate the role of AoA effects in orthography, we had previously exposed the model to approximately half of the vocabulary during pre-literacy training on the oral language tasks, so that observable effects of AoA could be related to reading exposure (see Table 1). However, the remaining half of the words that were not experienced prior to literacy should, if our predictions are correct, be processed along different pathways, and show a distinct psycholinguistic signature in the model's performance. Future simulations could implement AoA effects prior to as well as post literacy onset to examine more fine-grained effects of AoA before and after literacy. The disadvantage of simulating AoA in this way is then that the contribution of reading experience cannot be isolated from the effect of oral language experience, as is determinable in our current simulations.

For the model, words learned orally prior to literacy are those words with reading age between one and 13. Words learned in their entirety post literacy are those with reading age 14. Hence, if this prior learning of words orally (i.e., prior to literacy) affects the model's operation, then we should see a discontinuity in the effect of AoA in the model's performance between reading age 13 and 14, i.e., a discontinuity in the effect of AoA. This can be tested in two ways: either through multiple regression of an additional factor – literacy onset – in predicting the model's performance for word naming and lexical decision tasks; alternatively, this can be tested in terms of detecting discontinuities in the effect of AoA in the model using splines. Splines analysis determines

whether adding a discontinuity into the linear fit of a predictor variable to the dependent variable significantly improves the fit to the data. Baayen, Feldman, and Schreuder (2006) demonstrated this method for a range of psycholinguistic variables. New, Ferrand, Pallier, and Brysbaert (2006) also reported analyses of discontinuities in the length effect for word naming. However, neither of these studies reported analyses of age of acquisition. We report both methods of analysis below.

If the employment of different pathways in the model's performance is borne out for words acquired pre- versus post-literacy, then testing this behaviourally can be done with greater distinction using AoA measures that apply to pre- and post-literacy language development, which combine the consequences of oral and written experience of words. Thus, in investigating behavioural studies of use of different pathways, the signature of AoA will be slightly different but should still highlight a qualitative difference of the influence of AoA pre- versus post-literacy. We return to these behavioural analyses later.

Multiple regression analyses of literacy onset in the model

To examine the effect of literacy onset on the model's performance, hierarchical regression analyses were conducted. An additional variable, literacy onset, was created in which 2738 words learned orally pre-literacy onset were coded zero, and the remainder of words learned only post-literacy were coded one. At regression step 1, all previously used psycholinguistic variables were entered, then at step 2, literacy onset was included. If processing changes for words acquired orally prior to literacy onset compared to words acquired post-literacy onset, then the effect of AoA at the point of literacy onset should change, as an index of the involvement of semantics – reflected in a significant effect of literacy onset. If AoA is a continuous variable – such that there is a linear effect of AoA that applies for all reading ages including pre- and post-literacy onset, then there should be no additional variance associated with the literacy onset variable. The results for word naming and lexical decision are shown in Table 7.

The results showed the effect of literacy onset was a significant predictor of change in the model's performance. At the point of literacy onset, the regression gradient for the AoA effect was not sufficient to account for the variance associated with responses to words learned post-literacy onset. Words acquired pre-literacy and words acquired post-literacy were thus processed differently in the model. This effect was substantially larger for the model's lexical decision than word naming responses, consistent with suggestions that processing for pre-literacy acquired words used the indirect route from orthography to semantics via phonology, whereas the post-literacy acquired words used the direct orthography to semantics route.

We conducted an additional analysis using splines for AoA, to determine whether the discontinuity in AoA between age 13 and 14 was

Table 7
Results from two-step regression analyses for the exploration of literacy onset in predicting both word naming and lexical decision model performance.

		Word Naming	Lexical Decision
		Standardised β	Standardised β
Step 1	CF	−0.184***	−0.258***
	OrthN	−0.255**	0.041**
	Cons	−0.246***	−0.011
	Len	−0.069***	−0.111***
	Conc	−0.019	−0.012
	NoF	0.035**	0.029*
	SemN	0.0005	−0.233***
	AoA	0.192***	0.365***
Step 2	Literacy onset	0.228***	0.570***
		$\Delta R^2 = 0.37\%$	$\Delta R^2 = 2.27\%$

*** $p < .001$, ** $p < .01$, * $p < .05$.

significant. In line with the regression analyses of literacy onset, both word naming and lexical decision demonstrated significant discontinuities in AoA between age 13 and 14, $F(1, 5204) = 24.72$, $p < .001$, $F(1, 5226) = 181.32$, $p < .001$, respectively, with a larger effect for lexical decision.

Division of labour in the model for reading words acquired orally pre- and post-literacy

Several computational studies on reading have utilised a lesioning technique to explore the relevant contribution from different pathways to the activation of either semantics or phonology (Chang, Welbourne, & Lee, 2016; Harm & Seidenberg, 2004; Welbourne, Woollams, Crisp, & Lambon Ralph, 2011). For instance, Harm and Seidenberg (2004) demonstrated that for word comprehension, the indirect pathway from orthography to semantics via phonology was utilised more for generating semantics compared to the direct pathway from orthography to semantics, whereas for word naming, the direct orthography to phonology pathway was dominant, with only a small contribution via semantics. We examined the use of distinctive pathways for reading of words acquired orally pre- and post-literacy onset in the present developmental model using a similar method.

To obtain the contribution made by the orthography-to-phonology pathway (OP), we first computed the percent-correct score at the phonological layer after all the links from the semantics-to-phonology (SP) pathway to phonological units were removed. Similarly, for the contribution of the SP pathway, the percent-correct score at the phonological layer was computed after all the links from OP pathways to phonological units were removed. The relative contributions from OP and SP pathways to phonology were determined by calculating their proportional correct scores, and then this was used as an index of division of labour. Similar procedures were applied to calculate the contribution to semantics from the orthography-to-semantics (OS) and phonology-to-semantics (PS) pathways.

We tested the processing of words acquired orally pre- and post-literacy onset in the model in terms of division of labour between different pathways to the activation of phonology and semantics. ANOVA analyses were conducted, and the resulting patterns are shown in Fig. 7. For phonology, both pre-literacy and post-literacy words were processed mainly via the direct OP pathway relative to the indirect SP pathway, $F(1, 9) = 92,993$, $p < .001$. Furthermore, pre-literacy words also utilised the SP pathway to a small extent, but post-literacy words did not, $F(1, 9) = 101.3$, $p < .001$. For semantics, there is a clear differential pattern for pre-literacy and post-literacy words: pre-literacy words relied strongly on the indirect PS pathway to access semantics while post-literacy words utilised more equally the PS and OS pathways. The difference in reliance between the PS and OS pathways for post-literacy words was significantly smaller than pre-literacy words, $F(1, 9) = 137.4$, $p < .001$. Note that post-literacy words also relied more on the indirect PS pathway than the direct OS pathway.

During reading development the model's training was interleaved between oral language and reading trials. The division of labour results suggest that for post-literacy words the mappings from phonology to semantics are easier to acquire compared to the mappings from orthography to semantics, with the model able to focus learning on newly acquired words, with previously learned words resulting in low errors along these pathways. For mappings from orthography, both pre- and post-literacy words utilise high-fidelity representations within phonology and semantics for known words in acquiring novel words. For mappings from orthography, all mappings have to be acquired initially which means that learning along the pathways is not initially prioritised for any class of words. However, the efficiency of the mappings between phonology and semantics remains less refined than that for pre-literacy words.
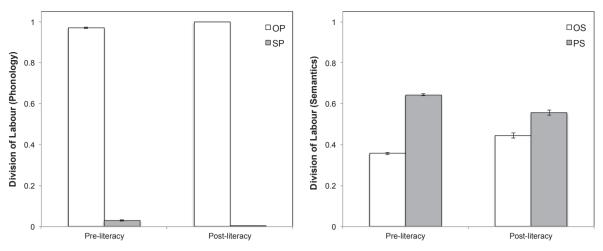
Fig. 7. The division of labour between different reading pathways for processing of pre- and post-literacy words.

Behavioural observations of literacy onset on reading

The model results, therefore, predicted that the linear effect of AoA would change at the point of literacy onset, that this would be most prominent for lexical decision, and that the continuing effect of this alternative use of pathways in the reading system ought to be observable in the behaviour of adult readers. We next tested these predictions using a dataset of behavioural responses to single words for word naming and lexical decision tasks. As noted above, due to the graded learning of words orally prior to literacy onset, reflected in behavioural AoA measures, we analysed the behavioural results using splines analysis to test whether there is a discontinuity in the regression of AoA against adults' word naming and lexical decision response times.

Data preparation

We analysed both word naming and lexical decision data from the English Lexicon Project (Balota et al., 2007), which represents a mega-study of word naming and lexical decision reaction times on 40,000 words averaged over responses from a subset of 1200 adult participants. For comparison with the computational modelling results, we focused on monosyllabic words. Unlike the simulation environment, the exact age of literacy onset in the behavioural data is unknown – as some words could still be learned from oral exposure even when formal literacy training begins. Thus, we determined the point at which adding a discontinuity into the linear fit of AoA resulted in the largest reduction in residual variance.

We tested the same set of psycholinguistic predictors as for the computational modelling analyses of word naming and lexical decision responses. Frequency was taken from the SUBTLEX-US corpus (Brysbaert & New, 2009). Orthographic neighbourhood size, length, and consistency were computed in the same way as for the modelling analysis. Concreteness was, as for the modelling analysis, taken from Brysbaert et al. (2014). Age of acquisition was taken from Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012), based on subjective ratings. For both word naming and lexical decision data, there were 1338 monosyllabic words included with all psycholinguistic predictors available. Descriptive statistics of the words are shown in Table 8.

Analysis

We analysed whether there were discontinuities in the fit of age of acquisition to both word naming and lexical decision response times separately, in terms of a change in the gradient of the linear fit. To do this, we determined whether adding a knot with degree 1 (so a change in the gradient of the linear fit of age of acquisition) significantly

Table 8
Descriptive statistics for the words included from the English Lexicon project in the analysis.

Variable	Mean	SD	Range
Age of Acquisition	7.64	2.70	2.22, 17.50
Frequency (ZipfSUBTLEX)	2.60	0.95	0.30, 6.33
Length	4.31	0.88	1, 8
Orthographic Neighbours	6.74	5.04	0, 24
Consistency	0.90	0.21	0.04, 1
Concreteness	3.04	1.42	1, 5

improved the model fit to predict the response times. We iteratively adjusted the position of the knot with respect to age of acquisition to find the point at which the fit improved most.

Results

Word naming

Fig. 8 (left panel) shows the sum of the squared residuals for the multiple linear regression model with a knot placed at age of acquisition values from age 2.2 to 18.6 years in intervals of 0.1 years. The minimum value was achieved at age of acquisition 15.9. Table 9 shows the multiple linear regression analysis with the knot for age of acquisition placed at age 15.9, adjusted $R^2 = 0.266$, F(7, 1330) = 70.27, $p < .001$. From age 2.2 to 15.9 years (indicated in Table 9 as *Age of Acquisition,* (1), the increase in response times per year was 3.4 ms per year (SE = 0.6 ms). From age 15.9 to 18.6 (indicated in the Table as *Age of Acquisition,* (2), the increase in response times per year was 46.5 ms per year (SE = 16.3 ms). We further tested whether this nonlinearity in the fit of AoA to the word naming response times accounted for significant variance in addition to a linear fit of AoA. However, model comparisons did not show a significant improvement for including the discontinuity, $F(1, 1330) = 2.748$, $p = 0.0976$.[1] This indicated that,

---

[1] We also tested restricted cubic splines which determine cubic nonlinearities in data, without distorting the endpoints of the data range (Harrell, 2015), and compared regression models with versus without the non-linearity. The results confirmed the linear effect of AoA was significant, $F(2, 1330) = 17.61$, $p < .001$, but the nonlinear effect was not significant, $F(1, 1330) = 1.31$, $p = .253$. For the lexical decision data, restricted cubic splines showed that the linear effect of AoA was significant, $F(2, 1330) = 92.51$, $p < .001$, and the nonlinear effect of AoA improved fit further, $F(1, 1330) = 11.05$, $p = .0009$. Note that restricted cubic splines are not appropriate for modelling the computational simulation data because the discontinuity is close to the end of the data range (between reading age 13 and 14), and so the discontinuity was best fitted with a single knot, as was conducted in the analysis of the computational
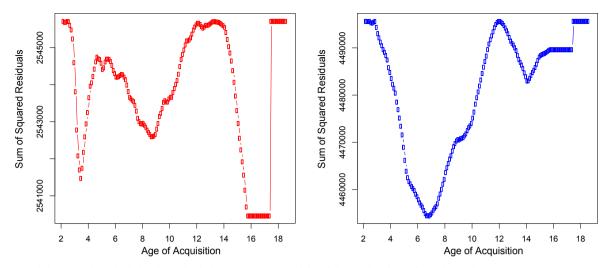
Fig. 8. The sum of the squared residuals for the multiple linear regression model with knots placed at values of age of acquisition from age 2.2 to 18.6 years in intervals of 0.1 years for word naming (left panel) and lexical decision (right panel) behavioural data.

Table 9
Multiple linear regression analyses of word naming response times for English Lexicon project monosyllabic words with age of acquisition spline degree 1, with a knot at age 15.9 years for word naming and a knot at age 6.8 years for lexical decision. Age of Acquisition 1 indicates the beta value for the regression from age 2.2 years to the knot, and Age of Acquisition 2 indicates the beta value for the regression from the knot to age 18.6 years.

	Word Naming	Lexical Decision
	Standardised β	Standardised β
Intercept	604.0***	755.1***
Age of Acquisition 1	5.64***	2.31*
Age of Acquisition 2	2.85**	13.04***
Frequency (ZipfSUBTLEX)	−5.45***	−12.88***
Length	7.03***	−3.62***
Orthographic Neighbours	−3.96***	−2.98**
Consistency	−4.48***	−2.41*
Concreteness	2.08*	2.28*

*** $p < .001$, ** $p < .01$, * $p < .05$.

though there was a perceived discontinuity in naming times, this was likely due to noise in the naming responses for later acquired words.

Lexical decision

Similar procedures were applied to the lexical decision data. The residuals were reduced most for a knot included at age 6.8 years (see Fig. 8, right panel). Table 9 shows the linear regression analysis for lexical decision with this knot at year 6.8 included, adjusted $R^2 = 0.467$, F(7, 1330) = 166.7, $p < .001$. From age 2.2 to 6.8 years, the increase in response times per year was still significant (4.4 ms per year, SE = 1.9 ms), but from age 6.8 to 18.6 the increase in response times per year increased more steeply (13.6 ms per year, SE = 1.0 ms). For testing whether the nonlinearity in the fit of AoA to lexical decision response times was significant, we compared the variance explained by nonlinear versus linear fits of AoA, confirming a significant discontinuity for lexical decision, F(1, 1330) = 12.30, $p = .0005$. This means that Age of Acquisition demonstrated an increasing effect after the age of 6.8 years. The fits of the splines to word naming and lexical decision data are shown in Fig. 9.

The results demonstrate the predicted discontinuity in the effect of AoA on lexical decision responses, around a time in development – age

(*footnote continued*)
model of literacy onset, above.

6.8 years – when children begin to learn novel words from text rather than from oral experience. This increasing gradient of AoA after age 6.8 years is consistent with the modelling data that show an enhanced role of direct orthography to semantics mappings for later-acquired than earlier-acquired words. Words learned prior to literacy onset are able to exploit the indirect orthography to semantics via phonology pathway for generating meanings.

General discussion

This paper aimed to explore unfolding experience in language processing by examining the AoA effects in a large-scale computational model of reading including orthographic, phonological and semantic representations with the incorporation of cumulative exposure to words whilst learning to read (Monaghan & Ellis, 2010). Despite introducing this complexity of staggered introduction of words during training to simulate gradual, incremental reading acquisition, the model was able to produce correct phonological and semantic patterns for naming of words given by their meanings, oral comprehension, word and non-word naming and lexical decision tasks. Multiple regression analyses on the model's performance demonstrated that the model was able to account for a range of standard reading effects including cumulative frequency, orthographic neighbourhood size, consistency, concreteness and the interactions between cumulative frequency and consistency, and AoA and consistency.

The simulation of gradual experience in presenting words to the model resulted in substantial advantages in simulating more subtle effects of psycholinguistic variables in reading. In particular, the results showed that AoA accounted for reliable, variance in both word naming and lexical decision, when other potentially confounding variables including cumulative frequency and concreteness had been considered. The attribution of AoA effects to the consequences of cumulative and incremental learning was further confirmed by additional simulations with randomised AoA. In addition, when all other lexical semantic variables were included into the regression analyses, there was no longer an effect of concreteness even though AoA effects remained, which is consistent with recent mega studies on lexical decision times in Dutch (Brysbaert, Stevens, Mandera, & Keuleers, 2016). Collectively, the regression results are largely consistent with the findings of previous regression analyses for both behavioural (Cortese & Khanna, 2007) and smaller-scale computational (Monaghan & Ellis, 2010) studies.

Further analysis of the model's functioning predicted that AoA would relate to differing use of the reading system, with words acquired
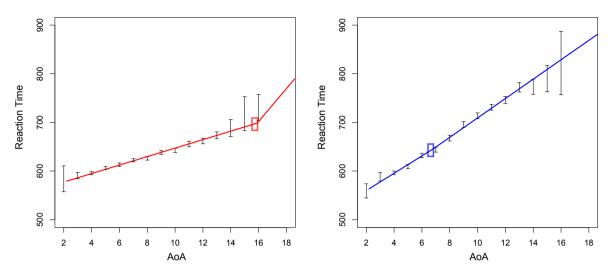
Fig. 9. The fits of the splines to word naming (left panel) and lexical decision (right panel) behavioural data.

before versus after literacy onset having distinctive access to meaning and sound representations, and these predictions were confirmed by an analysis of behaviour from the English Lexicon project.

Theories of AoA and their realisation in the model

Studies with AoA have provided different accounts for the locus of AoA (see Juhasz, 2005, for a review). The representation theory argues that the effect of AoA is due to differences in fidelity of representations of early versus late acquired words (e.g., Brysbaert et al., 2000), or differences in richness of representation (e.g., Steyvers & Tenenbaum, 2005). Alternatively, the mapping theory postulates the AoA effect results from the change in neuroplasticity in the system which would, with development, reduce the dynamic ability of the model to construct mappings between representations (e.g., Ellis & Lambon Ralph, 2000). An integrated view of AoA suggests that both representations and mappings could influence processing and contribute to AoA effects (Menenti & Burani, 2007).

In the literature, several connectionist models of word reading have provided evidence in support for the mapping theory (Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006; Monaghan & Ellis, 2010; Zevin & Seidenberg, 2002, 2004). However, none of those models have included more than a single processing route in the reading system. This prevented these modelling approaches from investigating the possibility of multiple potential sources of AoA and their contributions to the operation of the reading system. In the present study, the inclusion of the semantic system allowed the model to address such a possibility and to simulate differential AoA effects in word naming and lexical decision tasks, and to uncover fundamental differences in how the model reads words that are acquired pre-literacy and post-literacy.

As shown in Table 3, there were unique AoA effects that could be attributed to both the incremental learning of mappings between representations for word naming and lexical decision but also variance associated with AoA that was contained in differences in the semantic representations in terms of the density of their overlapping features. In addition, we predicted that the AoA effect would be moderated by consistency in word naming as an indicator of the locus of the effect of AoA in the mappings, and this prediction is supported by behavioural data reported by Monaghan and Ellis (2002) where the AoA effect is stronger for low consistency words than for high consistency words in a word naming task. As most words have typical spelling-to-sound mappings in English, later-learned regular words can be accommodated around pre-existing mappings without requiring substantial change in the connections. However, this would not be the case for later-learned words with atypical spelling-to-sound mappings which require

reorganisation of mappings, which cannot be so effectively accomplished when the connections are low in plasticity. The model's behaviour showed a similar interaction of AoA by consistency (Fig. 5) to that found in the behavioural data.

Both the representational and the mapping theories of AoA predict a larger effect of AoA when semantics is involved, and this effect was also reflected in the model. As mappings between orthography and semantics are arbitrary in nature, there is substantial cost for later-learned words to be integrated into pre-existing mappings, as learning each pattern cannot inherit mappings that are already in place between written and meaning forms (Lambon Ralph & Ehsan, 2006; Zevin & Seidenberg, 2002, 2004). In the model, the effect size of AoA was larger for lexical decision than word naming (Juhasz, 2005), see Table 3.

Taken together, these results demonstrated that AoA could not be solely determined by semantic representations, but also was present in mappings between representations. Consequently, the present simulation results provide evidence that a developmental model of reading results in AoA effects arising from multiple sources. Experience, then, affects cognitive processing in ways consistent with both representational richness and plasticity of mapping accounts of how early versus late experience influences behaviour.

Differential processing for words learned pre-literacy and post-literacy

The model's performance suggested that literacy onset would have an impact on the use of the architecture of the reading system. This was confirmed by the analysis of division of labour along alternative routes in the model. In particular, the model predicted differential AoA effects for words learned prior to literacy onset and after, particularly for tasks involving semantics, such as lexical decision or word comprehension. This was because the model was able to exploit the pre-existing mappings between phonology and semantics for words learned prior to literacy onset, enabling the indirect orthography to semantics route via phonology to be used for reading pre-literacy acquired words. For words acquired after literacy onset, the model predicted a greater effect of AoA, as the direct arbitrary mapping between orthography and semantics would be relied on more heavily for processing.

In the model, both word naming and lexical decision were affected by literacy onset, but the effect was substantially greater for lexical decision. In the behavioural results, there was also discovered a discontinuity in the influence of AoA on both word naming and lexical decision. However, the discontinuity was greater for lexical decision, as predicted, and occurred at a point consistent with age of literacy onset for children learning to read. Thus, the model's predictions about vestigial traces of the history of learning are observed in the adult's reading

system. This suggests that, for different words, reading advances along distinct pathways in the reading architecture. The model and the behavioural analysis of literacy onset suggests that words vary according to whether they use direct or indirect pathways in mapping between representations.

For the behavioural analysis of AoA in word naming, the discontinuity seems to occur at a relatively late point. It may be that, at a later age, reading for meaning might become more important for children compared to reading fluency, resulting in the increasing use of an indirect mapping from orthography to phonology via semantics which would increase the role of AoA in word naming. However, the amount of variability in the range of response times for word naming in this later acquired set, as indicated by the lack of significant effect of the discontinuity in the behavioural word naming data, may instead indicate that the word naming discontinuity is slight, or even non-existent.

However, for lexical decision, or other tasks involving activation of semantic representations, the role of literacy onset is quite different. When prior knowledge about phonological and semantic associations is available, as it is for pre-literacy acquired words, then an indirect route is likely to be involved in mapping from orthographic to semantic representations via phonology. For words learned post-literacy, this prior knowledge is not available, and so the reading system has to proceed via generating either a new mapping from orthography to semantics, or a new mapping from phonology to semantics. Thus, a distinct pattern of response is likely to be observed for lexical decision of pre- and post-literacy words at an early age.

A potential limitation of our approach is that we have focused on the modelling of orthographic AoA effects. As the WFG (Zeno et al., 1995) reports age-appropriate reading materials from age five upwards, it is not possible to use this same database to differentiate children's acquisition of words orally prior to literacy onset. However, the model was still able to raise predictions about different AoA effects before versus after literacy which were also observable in mega-studies of word naming and lexical decision responses. Isolating AoA effects to the orthographic training in the model ensured that we could relate the model's experience to chronological effects in reading development, but the general approach is consistent with introducing and investigating AoA effects in oral language training. In this case, we predict that oral AoA and orthographic AoA effects would both exert an influence on the model's performance: the oral AoA effects influencing the phonology-semantics mappings, and the orthographic AoA effects continuing to be observed in the use of direct and indirect pathways between orthography and phonology and orthography and semantics.

Taken together, the findings from both modelling and behavioural consistently demonstrate that even though literacy onset was several years before the participants in the reading studies are tested, the consequences of literacy onset are still observable in reading behaviour. We have shown that literacy onset changes the use that the reader makes of the language system, and this differential use of the system survives to be observed in behavioural responses even after decades of reading practice.

Conclusion

Taking a developmental modelling approach to reading has thus provided a range of benefits for understanding cognition of the reading system. The model provides a test for how AoA effects are realised in the language processing system, showing that both the representational and the mapping account of AoA are consistent with a model that learns to map between representations when those representations are presented incrementally as they are for children learning to read.

The model has implications more broadly for how models can provide insight into the role of development in understanding the mature cognitive processing system. A full understanding of cognitive processing is not contained in observations of the cognitive system's

current environment and response to that environment. Rather, we have shown it requires a consideration of how the individual's learning history impacts on representation and mapping between representations. Our example of developmental trajectory in a model, illuminated through an exploration of reading development, shows how the individual's experience of the environment can provide a fuller understanding for how the cognitive system reacts and adapts to this chronological experience.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2019.05.003.

References

Adelman, J. S., & Brown, G. D. A. (2008). Methods of testing and diagnosing model error: Dual and single route cascaded models of reading aloud. *Journal of Memory and Language, 59*(4), 524–544. https://doi.org/10.1016/j.jml.2007.11.008.

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814–823. https://doi.org/10.1111/j.1467-9280.2006.01787.x.

Adelman, J. S., Sabatos-DeVito, M. G., Marquis, S. J., & Estes, Z. (2014). Individual differences in reading aloud: A mega-study, item effects, and some models. *Cognitive Psychology, 68*, 113–160. https://doi.org/10.1016/j.cogpsych.2013.11.001.

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers, 36*(1), 140–155. https://doi.org/10.3758/bf03195559.

Anderson, K. L., & Cottrell, G. W. (2001). Age of acquisition and connectionist networks. *Proceedings of the 23rd annual meeting of the cognitive science society* (pp. 27–32). Mahwah, N.J: Lawrence Erlbaum.

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition, 10*(6), 565–575. https://doi.org/10.3758/bf03202439.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 234–254. https://doi.org/10.1037/0278-7393.18.2.234.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language, 55*(2), 290–313. https://doi.org/10.1016/j.jml.2006.03.008.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445–459. https://doi.org/10.3758/bf03193014.

Barry, C., Johnston, R. A., & Wood, R. F. (2006). Effects of age of acquisition, age, and repetition priming on object naming. *Visual Cognition, 13*(7–8), 911–927. https://doi.org/10.1080/13506280544000101.

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *The Quarterly Journal of Experimental Psychology Section A, 50*(3), 560–585. https://doi.org/10.1080/783663595.

Bates, E., Bretherton, I., Snyder, L., Beeghly, M., Shore, C., McNew, S., et al. (1988). *From first words to grammar: Individual differences and dissociable mechanisms.* New York, NY, US: Cambridge University Press.

Bates, E., Burani, C., D'Amico, S., & Barca, L. (2001). Word reading and picture naming in Italian. *Memory & Cognition, 29*(7), 986–999. https://doi.org/10.3758/bf03195761.

Belsky, J., Barnes, J., & Melhuish, E. C. (Eds.). (2007). *The national evaluation of sure start: Does area-based early intervention work?.* Bristol: Bristol University Press.

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*(4), 456–476. https://doi.org/10.1016/j.jml.2004.02.001.

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology, 93*(1), 89–114. https://doi.org/10.1348/000712602162463.

Bradley, V., Davies, R., Parris, B., Su, I. F., & Weekes, B. S. (2006). Age of acquisition effects on action naming in progressive fluent aphasia. *Brain and Language, 99*(1),

128–129. https://doi.org/10.1016/j.bandl.2006.06.073.

Brown, G. D. A., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition, 15*(3), 208–216. https://doi.org/10.3758/bf03197718.

Brysbaert, M. (2017). Age of acquisition ratings score better on criterion validity than frequency trajectory or ratings "corrected" for frequency. *Quarterly Journal of Experimental Psychology, 70*(7), 1129–1139. https://doi.org/10.1080/17470218. 2016.1172097.

Brysbaert, M., & Ellis, A. W. (2016). Aphasia and age of acquisition: Are early-learned words more resilient? *Aphasiology, 30*(11), 1240–1263. https://doi.org/10.1080/02687038.2015.1106439.

Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition, 13*(7–8), 992–1011. https://doi.org/10.1080/13506280544000165.

Brysbaert, M., Lange, M., & van Wijnendaele, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology, 12*(1), 65–85. https://doi.org/10.1080/095414400382208.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990. https://doi.org/10.3758/brm.41.4.977.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance, 42*(3), 441–458. https://doi.org/10.1037/xhp0000159.

Brysbaert, M., van Wijnendaele, I., & Deyne, S. D. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica, 104*(2), 215–226. https://doi.org/10.1016/S0001-6918(00)00021-4.

Brysbaert, M., & Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5.

Butler, B., & Hains, S. (1979). Individual differences in word recognition latency. *Memory & Cognition, 7*(2), 68–76. https://doi.org/10.3758/bf03197587.

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology, 25*(1), 85–95. https://doi.org/10.1080/14640747308440325.

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 psycholinguistic variables, the English data from Morrison et al. (1997), and naming latencies. *European Journal of Cognitive Psychology, 15*(2), 209–245. https://doi.org/10.1080/09541440244000076.

Chang, Y.-N., Lambon Ralph, M. A., Furber, S., & Welbourne, S. (2013). Modelling graded semantic effects in lexical decision. In Proceedings of the 35th annual meeting of the cognitive science society, Berlin, Germany (pp. 310–315).

Chang, Y.-N., Furber, S., & Welbourne, S. (2012). "Serial" effects in parallel models of reading. *Cognitive Psychology, 64*(4), 267–291. https://doi.org/10.1016/j.cogpsych. 2012.01.002.

Chang, Y.-N., & Monaghan, P. (2019). Quantity and diversity of preliteracy language exposure both affect literacy development: Evidence from a computational model of reading. *Scientific Studies of Reading, 23*(3), 235–253. https://doi.org/10.1080/10888438.2018.1529177.

Chang, Y.-N., Welbourne, S., & Lee, C.-Y. (2016). Exploring orthographic neighborhood size effects in a computational model of Chinese character naming. *Cognitive Psychology, 91*, 1–23. https://doi.org/10.1016/j.cogpsych.2016.09.001.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.). *Attention and Performance VI* (pp. 535–555). Hillsdale, NJ: Lawrence Erlbaum Associates.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204.

Coltheart, V., Laxon, V., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology, 79*(1), 1–12. https://doi.org/10.1111/j.2044-8295.1988.tb02270.x.

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology, 60*(8), 1072–1082. https://doi.org/10.1080/17470210701315467.

Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology, 66*(5), 946–972. https://doi.org/10.1080/17470218.2012.722660.

Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, & Computers, 31*(4), 650–658. https://doi.org/10.3758/bf03200741.

Davies, R., Arnell, R., Birchenough, J. M. H., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(8), 1298–1338. https://doi.org/10.1037/xlm0000366.

Davies, R., Wilson, M., Cuetos, F., & Burani, C. (2014). Reading in Spanish and Italian: Effects of age of acquisition in transparent orthographies? *Quarterly Journal of Experimental Psychology, 67*(9), 1808–1825. https://doi.org/10.1080/17470218. 2013.872155.

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*(5), 1103–1123. https://doi.org/10.1037/0278-7393.26.5.1103.

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(2), 515–523. https://doi.org/10.1037/0278-7393.24.2.515.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179–211. https://doi.org/10.1016/0364-0213(90)90002-E.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*(1), 71–99. https://doi.org/10.1016/0010-0277(93)90058-4.

Gerhand, S., & Barry, C. (1999). Age-of-acquisition and frequency effects in speeded word naming. *Cognition, 73*(2), B27–B36. https://doi.org/10.1016/S0010-0277(99)00052-9.

Ghyselinck, M., Custers, R., & Brysbaert, M. (2004). The effect of age of acquisition in visual word processing: Further evidence for the semantic hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 550–554. https://doi.org/10.1037/0278-7393.30.2.550.

Ghyselinck, M., Lewis, M. B., & Brysbaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. *Acta Psychologica, 115*(1), 43–67. https://doi.org/10.1016/j.actpsy.2003.11.002.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12*(4), 395–427. https://doi.org/10.3758/bf03201693.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*(4), 674–691. https://doi.org/10.1037/0096-1523.5.4.674.

Grondin, R., Lupker, S. J., & McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *Journal of Memory and Language, 60*(1), 1–19. https://doi.org/10.1016/j.jml.2008.09.001.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*(3), 491–528. https://doi.org/10.1037/0033-295X.106.3.491.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111*(3), 662–720. https://doi.org/10.1037/0033-295X.111.3.662.

Harold, G., Acquah, D., Sellers, R., & Chowdry, H. (2016). What Works to Enhance Inter-Parental Relationships and Improve Outcomes for Children? Department for Work and Pensions. Early Intervention Foundation Report. UK Government: Department for Work and Pensions. http://www.eif.org.uk/publication/what-works-to-enhance-inter-parental-relationships-and-improve-outcomes-for-children-3. Retrieved 22nd August 2016.

Harrell, F. E. (2015). *Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed.). New York: Springer.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children.* Baltimore: Brookes Publishing Co.

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin, 133*(4), 638–650. https://doi.org/10.1037/0033-2909.133.4.638.

Hino, Y., & Lupker, S. J. (2000). Effects of word frequency and spelling–sound regularity in naming with and without preceding lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 166–183. https://doi.org/10.1037/0096-1523.26.1.166.

Hodgson, C., & Ellis, A. W. (1998). Last in, First to Go: Age of acquisition and naming in the elderly. *Brain and Language, 64*(1), 146–163. https://doi.org/10.1006/brln.1998.1960.

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods, 45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x.

Holmes, S. J., Jane Fitch, F., & Ellis, A. W. (2006). Age of acquisition affects object recognition and naming in patients with Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 28*(6), 1010–1022. https://doi.org/10.1080/13803390591004392.

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language, 103*, 114–126. https://doi.org/10.1016/j.jml.2018.08.005.

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read: Combining intervention studies with mediation analyses. *Psychological Science, 23*(6), 572–577. https://doi.org/10.1177/0956797611435921.

Izura, C., & Ellis, A. W. (2002). Age of acquisition effects in word recognition and production in first and second languages. *Psicológica, 23*(2), 245–281.

Izura, C., & Ellis, A. W. (2004). Age of acquisition effects in translation judgement tasks. *Journal of Memory and Language, 50*(2), 165–181. https://doi.org/10.1016/j.jml.2003.09.004.

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition, 13*(7–8), 789–845. https://doi.org/10.1080/13506280544000066.

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin, 131*(5), 684–712. https://doi.org/10.1037/0033-2909.131.5.684.

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1312–1318. https://doi.org/10.1037/0278-7393.29.6.1312.

Kosmidis, M. H., Tsapkini, K., Folia, V., Vlahou, C. H., & Kiosseoglou, G. (2004). Semantic and phonological processing in illiteracy. *Journal of the International Neuropsychological Society, 10*(6), 818–827. https://doi.org/10.1017/

S1355617704106036.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4.

Lake, B. M., & Cottrell, G. W. (2005). Age of acquisition in facial identification: A connectionist approach. *Proceedings of the 27th cognitive science society conference* (pp. 1236–1341). Mahwah, NJ: Lawrence: Erlbaum.

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*(7–8), 928–948. https://doi.org/10.1080/13506280544000110.

Lewis, M. B., Gerhand, S., & Ellis, H. D. (2001). Re-evaluating age-of-acquisition effects: Are they simply cumulative-frequency effects? *Cognition, 78*(2), 189–205. https://doi.org/10.1016/S0010-0277(00)00117-7.

Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks, 17*(8), 1345–1362. https://doi.org/10.1016/j.neunet.2004.07.004.

Maki, W. S., McKinley, L. N., & Thompson, A. G. (2004). Semantic distance norms computed from an electronic dictionary (WordNet). *Behavior Research Methods, Instruments, & Computers, 36*(3), 421–431. https://doi.org/10.3758/bf03195590.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language, 92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001.

Marchman, V. A. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience, 5*(2), 215–234. https://doi.org/10.1162/jocn.1993.5.2.215.

Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*(2), 313–330.

Mareschal, D., Johnson, M. H., Sirois, S., Spratling, M., Thomas, M. S., & Westermann, G. (2007). *Neuroconstructivism: How the brain constructs cognition.* Oxford: Oxford University Press.

McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 14–24. https://doi.org/10.1037/0096-1523.13.1.14.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Vol. Ed.), *The psychology of learning and motivation: Vol. 24*, (pp. 109–165). New York: Academic Press.

Menenti, L., & Burani, C. (2007). What causes the effect of age of acquisition in lexical processing? *The Quarterly Journal of Experimental Psychology, 60*(5), 652–660. https://doi.org/10.1080/17470210601100126.

Meschyan, G., & Hernandez, A. (2002). Age of acquisition and word frequency: Determinants of object-naming speed and accuracy. *Memory & Cognition, 30*(2), 262–269. https://doi.org/10.3758/bf03195287.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography, 3*(4), 235–244. https://doi.org/10.1093/ijl/3.4.235.

Monaghan, J., & Ellis, A. W. (2002). What exactly interacts with spelling–sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 183–206. https://doi.org/10.1037/0278-7393.28.1.183.

Monaghan, P., Chang, Y.-N., Welbourne, S., & Brysbaert, M. (2017). Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language, 93*, 1–21. https://doi.org/10.1016/j.jml.2016.08.003.

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*(4), 506–525. https://doi.org/10.1016/j.jml.2010.08.003.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*(4), 323–331. https://doi.org/10.1016/0010-0277(79)90020-9.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 116–133. https://doi.org/10.1037/0278-7393.21.1.116.

Morrison, C. M., & Ellis, A. W. (1999). Age of acquisition, lexical processing and ageing: Changes across the lifespan. *Proceedings of the 21st annual meeting of the cognitive science society* (pp. 444–447). Mahwah, N.J: Erlbaum.

Morrison, C. M., Hirsh, K. W., Chappell, T., & Ellis, A. W. (2002). Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *European Journal of Cognitive Psychology, 14*(4), 435–459. https://doi.org/10.1080/09541440143000159.

Nation, K. (2009). Form-meaning links in the development of visual word recognition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1536), 3665–3674. https://doi.org/10.1098/rstb.2009.0119.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13*(1), 45–52. https://doi.org/10.3758/bf03193811.

Paap, K. R., & Noel, R. W. (1991). Dual-route models of print to sound: Still a good horse race. *Psychological Research, 53*(1), 13–24. https://doi.org/10.1007/bf00867328.

Pagliuca, G. & Monaghan, P. (2006). Dissociating word reading and lexical decision in neglect dyslexia: A connectionist account. In Proceedings of the 28th annual conference of the cognitive science society.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*(2), 273–315. https://doi.org/10.1037/0033-295X.114.2.273.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology, 61*(2), 106–151. https://doi.org/10.1016/j.cogpsych.2010.04.001.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes, 12*(5–6), 765–806. https://doi.org/10.1080/016909697386682.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56–115. https://doi.org/10.1037/0033-295X.103.1.56.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*(5), 377–500. https://doi.org/10.1080/02643299308253469.

Richardson, F. M., & Thomas, M. S. C. (2008). Critical periods and catastrophic interference effects in the development of self-organizing feature maps. *Developmental Science, 11*(3), 371–389. https://doi.org/10.1111/j.1467-7687.2008.00682.x.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83*(5), 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x.

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science, 323*(5916), 951–953. https://doi.org/10.1126/science.1167025.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523–568. https://doi.org/10.1037/0033-295X.96.4.523.

Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word-reading models at the item level: Matching the grain of theory and data. *Psychological Science, 9*(3), 234–237. https://doi.org/10.1111/1467-9280.00046.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23*(3), 383–404. https://doi.org/10.1016/S0022-5371(84)90270-6.

Sibley, D. A., Kello, C. T., & Seidenberg, M. S. (2009). Error, error everywhere: A look at megastudies of word reading. *Proceedings of the 31st annual conference of the cognitive science society* (pp. 1036–1041). Amsterdam, Netherlands: Cognitive Science Society.

Smith, M. A., Cottrell, G. W., & Anderson, K. L. (2001). The early word catches the weights. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.). *Advances in neural information systems 13* (pp. 52–57). Cambridge, MA: MIT Press.

Smith, A. C., Monaghan, P., & Huettig, F. (2014). Examining strains and symptoms of the 'Literacy Virus': The effects of orthographic transparency on phonological processing in a connectionist model of reading. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.). *Proceedings of the 36th annual meeting of the cognitive science society.* Austin, TX: Cognitive Science Society.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8*(6), 411–416. https://doi.org/10.1111/j.1467-9280.1997.tb00453.x.

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods, 38*(4), 598–605. https://doi.org/10.3758/bf03193891.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.). *Experimental cognitive psychology and its applications* (pp. 237–249). Washington, DC: American Psychological Association.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29*(1), 41–78. https://doi.org/10.1207/s15516709cog2901_3.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1140–1154. https://doi.org/10.1037/0278-7393.21.5.1140.

Strain, E., Patterson, K., & Seidenberg, M. S. (2002). Theories of word naming interact with spelling–sound consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 207–214. https://doi.org/10.1037/0278-7393.28.1.207.

Szwed, M., Ventura, P., Querido, L., Cohen, L., & Dehaene, S. (2012). Reading acquisition enhances an early visual process of contour integration. *Developmental Science, 15*(1), 139–149. https://doi.org/10.1111/j.1467-7687.2011.01102.x.

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language, 26*(6), 608–631. https://doi.org/10.1016/0749-596X(87)90105-7.

Taylor, J. S. H., Duff, F. J., Woollams, A. M., Monaghan, P., & Ricketts, J. (2015). How word meaning influences word reading. *Current Directions in Psychological Science, 24*(4), 322–328. https://doi.org/10.1177/0963721415574980.

The White House (2013). The White House summit on early childhood education. Retrieved 22 August 2016 https://www.whitehouse.gov/the-press-office/2014/12/10/fact-sheet-invest-us-white-house-summit-early-childhood-education.

Thomas, M. S. C., & Knowland, V. (2009). Sensitive periods in brain development – Implications for education policy (Vol. 2).

Thomas, M. S. C., & Johnson, M. H. (2008). New advances in understanding sensitive periods in brain development. *Current Directions in Psychological Science, 17*(1), 1–5. https://doi.org/10.1111/j.1467-8721.2008.00537.x.

Ueno, T., & Lambon Ralph, M. (2013). The roles of the "ventral" semantic and "dorsal" pathways in conduite d'approche: A neuroanatomically-constrained computational modeling investigation. *Frontiers in Human Neuroscience, 7*(422), 422. https://doi.org/10.3389/fnhum.2013.00422.

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal*

of Experimental Psychology, 67(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521.

van Loon-Vervoorn, W. A., van Ham, P., & van der Koppen, M. (1988). The importance of age of acquisition for imageability in word processing. In M. Dennis, J. Engelkamp, & J. T. E. Richardson (Eds.). Cognitive and neuropsychological approaches to mental imagery (pp. 99–107). Dordrecht: Martinus Nijhoff.

Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time-course and decision criteria. Memory & Cognition, 13(6), 557–572. https://doi.org/10.3758/bf03198326.

Welbourne, S. R., Woollams, A. M., Crisp, J., & Lambon Ralph, M. A. (2011). The role of plasticity-related functional reorganization in the explanation of central dyslexias. Cognitive Neuropsychology, 28(2), 65–108. https://doi.org/10.1080/02643294.2011.621937.

Westermann, G., Mareschal, D., Johnson, M. H., Sirois, S., Spratling, M. W., & Thomas, M. S. C. (2007). Neuroconstructivism. Developmental Science, 10(1), 75–83. https://doi.org/10.1111/j.1467-7687.2007.00567.x.

Whaley, C. P. (1978). Word—nonword classification time. Journal of Verbal Learning and Verbal Behavior, 17(2), 143–154. https://doi.org/10.1016/S0022-5371(78)90110-X.

Wilson, M. A., Ellis, A. W., & Burani, C. (2012). Age-of-acquisition affects word naming in Italian only when stress is irregular. Acta Psychologica, 139(3), 417–424. https://doi.org/10.1016/j.actpsy.2011.12.012.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. Journal of Experimental Psychology: Human Perception and Performance, 38(1), 53–79. https://doi.org/10.1037/a0024177.

Zeno, S. M., Ivens, S. H., Hillard, R. T., & Duvvuri, R. (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), 1–29. https://doi.org/10.1006/jmla.2001.2834.

Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. Memory & Cognition, 32(1), 31–38. https://doi.org/10.3758/bf03195818.