Spoken-word recognition in 2-year-olds: The tug of war between phonological and semantic activation

Janette Chow, Anne Aimola Davies and Kim Plunkett

Department of Experimental Psychology, University of Oxford

Abstract

Previous studies demonstrate that while toddlers can match words with their referent before the age of one, they only begin to extract phonologically- and semantically-related information from speech later in the second year. However, the order and manner in which this information is extracted remains unresolved. In two experiments, we adapted the adult four-picture Visual World Paradigm (VWP) (Huettig & McQueen, 2007) for toddler testing: toddlers hear a spoken word and see pictures that are phonologically-related, semantically-related or unrelated to the spoken-word. We demonstrate that, similar to adults, 24- to 30-month-old toddlers attended to a phonological match faster than a semantic match, and that these differences hold irrespective of whether the semantic match is taxonomic or thematic in character. Our findings suggest that language shapes toddlers' cognition by biasing their selective attention to relevant information in the visual world, which is likely to enhance the efficiency of mental activities, such as learning and making analogical inferences.

Keywords: visual world paradigm; lexical access; spoken-word recognition; lexicon; semantics; taxonomy; language development; eye tracking; visual attention; toddler

Author's note

J.C. designed and performed experiments, analysed the data, and wrote the manuscript; A.A.D. designed the experiments and critically reviewed the manuscript. K.P designed the experiments, and wrote and critically reviewed the manuscript.

The Authors declare that there is no conflict of interest.

We thank Mihaela Duta as the developer of the eye-tracking software *Presentmate* and gaze-processing codes, and Dan Lunn for his advice and help on implementing the statistical models. We also thank Nadja Althaus and Suzy Styles for their advice during the initial stage of the project, Amanda Griffins and Irina Lepadatu for participant recruitment, and all the toddler and parent participants who took part in the study.

Spoken-word recognition in 2-year-olds: The tug of war between phonological and semantic activation

Toddlers can match words with their referents before the age of one. Do toddlers simply associate a spoken label directly with its referent much as a pigeon learns that a light signals reward, or do they also access phonologically- and semantically-related information from the lexical-semantic system in a cascaded fashion as has been observed with adults? Numerous studies have demonstrated that toddlers, by the age of two, extract phonologically- and semantically-related information from speech. However, the order and manner in which this information is extracted remains unresolved. This study provides the missing piece of evidence by examining the time course of the online mental processes involved in spoken-word recognition in 24- to 30-month-old toddlers by adapting the adult Visual World Paradigm (VWP) for toddler testing. Furthermore, this study attempts to provide insights into the manner in which spoken-word recognition drives visual attention during toddlerhood, a critical period of language acquisition. In particular, how does the encoding of phonological and semantic relations affect the way toddlers direct their attention to objects in the world?

Over the past 20 years, the VWP has become a popular tool for the study of language processing in adults (see Huettig, Rommers, & Meyer, 2011, for a detailed review). One line of VWP research focuses on measuring the manner in which lexical-semantic representations are activated during spoken-word recognition. Cooper (1974) first demonstrated experimentally that adults would spontaneously and automatically fixate an object that was named or relevant to what was just said. In his studies, participants were presented with a grid of nine pictures, and heard a story. When participants heard the word ‘lion’, their eyes were drawn to the picture of the lion. Upon hearing the word ‘Africa’, they were also more likely to fixate the pictures of lion,

zebra and snake. These findings demonstrate that visual attention can be driven by the matching of auditory and visual stimuli, even when there is no explicit task or requirement for an overt response. The fact that participants fixated the related but unnamed pictures suggests that target-related representations were activated and retrieved from the lexical-semantic system, and matched with the current visual input, which in turn led to increased visual attention. Cooper's (1974) study provided concrete evidence that language mediates the listener's eye-movements, and that language-mediated eye-movements reflect mental activities of language processing.

Phonological Representations

The Cohort model of word recognition (Marslen-Wilson, 1987) proposes that the initial sounds of a word activate a set of candidate words in the lexicon that begin with the same phonemes. As the spoken word unfolds, candidate words that no longer match the speech input are eliminated until only one word remains. For example, upon hearing the first sound /d/ of the word 'dog', candidate words, such as 'dog', 'digger' and 'dress', are activated. Once the spoken vowel is processed, 'digger' and 'dress' can be discarded. The Cohort model assumes that lexical access is a strictly bottom-up process; top-down effects, such as context, only influence word selection among the candidate words. The TRACE model (McClelland & Elman, 1986) further embodies the proposal that spoken word recognition is not only incremental but also continuous, and therefore, words that match the sensory input in the middle or at the end of the signal, such as rhyme words, will also be activated (e.g., 'dog' will activate 'log'). Using the VWP, Allopenna, Magnuson, and Tanenhaus (1998) presented participants with a four-picture display including a target, cohort distractor (same onset consonant and first vowel as target), rhyme distractor (rhymes with target) and unrelated distractor. Upon hearing the target word, participants not only fixated the target picture, but also fixated the cohort distractor. Later in the

time course, participants also fixated the rhyme distractor more than the unrelated distractor. Their findings provide strong evidence that spoken-word recognition involves continuous incremental mapping of speech sounds and the automatic activation and elimination of competing phonological representations.

Semantic Representations

VWP studies have also provided evidence that spoken-word recognition leads to the automatic activation of semantically-related representations. Taxonomic and thematic relations are the two most studied types of semantic relations. Taxonomic relations refer to the classification or hierarchy of items. Items that are taxonomically-related tend to resemble each other, sharing perceptual, biological or functional properties (e.g., Medin & Smith, 1984; Smith & Medin, 1981). Taxonomic categories can be formed at different levels in the hierarchy, such as basic, subordinate and superordinate level categories. For example, labrador and collie belong to the category of dog; dog and horse belong to the category of mammals; mammals and birds belong to the category of animal. In contrast, thematic relations refer to the grouping of items co-occurring in the same space and time, often based on memory and personal experience. Rather than resembling each other, thematically-related items tend to complement each other. Amongst the most studied types of thematic relations are functional relations. These include at least two sub-types: instrument and script (Moss, Ostrin, Tyler, & Marslen-Wilson, 1995). The instrument relation describes word pairs in which one performs some action on the other. For example, shampoo is used on hair. The script relation describes the relationship between a context and items that occur or appear in the same time and space. For example, one finds sheep on a farm. Other commonly studied thematic relations include part-whole (e.g., coin is made of copper), causal-relation (e.g., wind propels a sail) and temporal-relation (e.g., breakfast takes place before

lunch). Often, thematically-related items are related in more than one way (Lin & Murphy, 2001). For example, chalk and blackboard are both instrument- and script-related, because one uses chalk to write on the blackboard, and you often find a chalk resting on the ledge of a blackboard.

Studies using the VWP have shown that upon hearing a target word (e.g., ‘dog’), the listener is more likely to fixate a taxonomically-related picture (e.g., pig) (Huettig & Altmann, 2005) or a thematically-related picture (e.g., bone) (Mirman & Graziano, 2012; Yee, Overton, & Thompson-Schill, 2009) than an unrelated distractor. These findings are consistent with localist models of semantic memory (e.g., Collins & Loftus, 1975) in which concepts are represented by nodes in a semantic network, and that concepts having a similar meaning share connections with each other within the network. When spoken-word or object recognition takes place, the activation of the referent node spreads to related nodes (i.e., semantically-related concepts), thereby underpinning language-mediated eye-movements in a VWP task. However, these results can also be taken as evidence for distributed models (e.g., Cree & McRae, 2003; Moss, Hare, Day, & Tyler, 1994; Tyler & Moss, 2001). Distributed or featured-based models argue that objects are categorised based on the degree of feature correlation or overlap. Correlation between taxonomically-related objects is achieved through the high number of shared-features (e.g., head and body in mammals). In contrast, the objects’ distinctive, uncorrelated features (e.g., zebras have stripes and deers have antlers) allow discrimination within a category (Cree & McRae, 2003; Tyler & Moss, 2001). Thematically-related objects can be correlated in terms of their co-occurrence in episodic memories or the co-occurrence of lexical forms (e.g., ‘shampoo’ and ‘hair’) in spoken utterances or written text (Moss et al., 1994; Plaut & Booth, 2000). In the VWP, fixation of taxonomically and thematically-related pictures is motivated by the activation of

shared features and co-occurrence associations with the target.

The Tug of War

Huetting and McQueen (2007) used the VWP to investigate the activation of lexical, semantic and visual-form representations during spoken-word recognition. In their Experiment 1, participants heard a sentence containing a target word, e.g., beaker¹ (Dutch: 'beker'), and saw a four-picture display containing a phonological distractor, semantic distractor, visual-form distractor and an unrelated distractor. The phonological distractor, e.g., beaver (Dutch: bever), had the same onset consonant as the target. The semantic distractor, e.g., fork (Dutch: vork), was either taxonomically- or thematically-related to and associated to the target. The visual-form distractor, e.g., bobbin (Dutch: klos) had the same shape as the target. The unrelated distractor, e.g., umbrella (Dutch: paraplu) was phonologically, semantically and visually dissimilar to the target. The authors found that participants were more likely to fixate the phonological distractor approximately 300–600ms after the onset of the target word, and the semantic and visual-form distractors approximately 400–1000ms after onset. These findings are consistent with the proposal that the activation of phonological, semantic and visual-form representations are separate processes during spoken-word recognition. Moreover, as speech is processed and recognised incrementally, the activation of phonological representations naturally precedes that of semantic and visual-form representations.

Locus of Word-Picture Matching

VWP studies often assume that language-mediated attention involves the matching of cascaded activation of stored knowledge between two routes—the audio-input (word) route and the visual-input (picture) route. In the audio-input route, perceiving the auditory signal leads to

¹ Huetting and McQueen (2007) translated 'beker' as 'beaker' in English, but it can also be translated as 'cup'.

cascaded activation of existing phonological representations (e.g., cohort candidates), followed by semantic (i.e., meaning) and visual-form representations. This is demonstrated by the findings of an earlier phonological preference and a later semantic and visual-form preference in adult VWP studies that involve spoken-word recognition (Huetting & McQueen, 2007). The visual-input route describes a cascade of activation in the reverse direction: Viewing a picture leads to cascaded activation of its visual-form representations, followed by semantic representations and may be its word form or phonological representations. It is generally assumed that participants' increased visual attention for a picture is the result of the matching cascaded activations in the audio-input and visual-input routes. However, the exact level in which activated information is matched in the two routes is currently still under debate.

The phonological hypothesis proposes that the 'phonological effect' is driven by matching of activations at the phonological level. That is, participants' increased visual attention towards the phonological distractor is driven by the matching of cascaded phonological activation from the audio-input route and the cascaded phonological activation from the visual-input. Similarly, the semantic hypothesis proposes that the 'semantic effect' is driven by the matching of semantic activations from the audio- and visual-input routes. Compared to cascaded semantic activation, the notion of cascaded phonological activation in the visual-input route is more controversial: Object recognition typically involves access to semantic information, but not necessarily its name. However, there is growing evidence for cascaded phonological activation in the visual-input route from both adult and toddler studies. In toddler priming studies, the presentation of a picture prime in silence can lead to phonological and phono-semantic priming effects (Mani & Plunkett, 2010, 2011; Mani, Durrant & Floccia, 2012), suggesting that toddlers extract the phonological codes from the visual input. In the adult literature, supporting evidence

comes from word production studies and visual search studies, which show that participants' response can be influenced by the presence of a phonologically-related distractor picture (e.g., Meyer, Belke, Telling & Humphreys, 2007; Meyer & Damian, 2007; Morsella & Miozzo, 2002, Navarette & Costa, 2005)². More recently, McQueen & Huettig (2014) show that spoken-word recognition is slowed down when adults are presented with an unnamed phonological picture prime. These findings suggest that some form of phonological code is available for each of the pictures in the display, particularly given that the viewers have had several seconds to view them.

Supporting evidence for the phonological and semantic hypotheses comes from VWP studies that use picture stimuli. Huettig and McQueen (2007) demonstrate (in Experiment 1) that when participants were given sufficient time to preview the display, participants demonstrated an earlier phonological preference and a later semantic and visual-form preference. In contrast, when participants did not have time to preview the display (in Experiment 2), they showed no preference for the phonological distractor. The authors argue that participants have insufficient time from display onset for cascaded phonological activation in the visual-input route to match the auditory input (see also De Groot et al., 2015 for similar findings with semantic and visual-form distractors).

The visual hypothesis proposes that the 'visual-form effect' is driven by the matching of visual-form activations (e.g., looking at a rope when they hear the word 'snake') (Dahan & Tanenhaus, 2005). The strong version of the visual hypothesis argues against the automatic activation of phonological or lexical information during passive picture viewing, and that a phonological effect can be driven by a visual match. For example, when one hears the /b/ sound of 'beaker', the phonological representations of cohort competitors, such as beaver, become

² But see Strijkers, Holcomb & Costa (2011) for a review of studies reporting opposite or null-effect findings.

activated, followed by the cascaded activation of the visual-form representations of beaver. A phonological effect can, therefore, be driven by the matching of the cascaded visual-form of beaver from the audio-input and the visual form of beaver from the visual-input. VWP studies using text stimuli provide evidence for the visual hypothesis: Salverda and Tanenhaus (2010) found that adult participants, who were given only 200ms preview of the text stimuli before hearing the spoken-word, showed a significant preference for a printed word that is orthographically more similar to the spoken-target word, compared to one that is orthographically less similar. In contrast, participants showed no significant preference for a printed word that has a greater phonological overlap with the spoken-target word, compared to one that has less phonological overlap. These findings provide evidence that the matching of spoken- and written-words takes place at the visual-form level.

Infant Inter-modal Preferential Looking Paradigm

The study of language-mediated visual attention is not novel in the infant literature, though until recently it has not been recognised as such; it has been measured using the inter-modal preferential looking (IPL) paradigm since the 1980s (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987). The rationale behind the adult VWP and infant IPL is the same—both paradigms measure eye-gazes that are interpreted as reflecting language processing. However, there are two main differences between the two paradigms. First, the design of many adult VWP studies involves target-absent displays, while the design of most infant IPL studies are target-present. In adult VWP studies, the aim is often to investigate visual attention towards non-target but relevant pictures, while the aim of infant IPL studies is often simply to investigate whether infants have learnt and/or understood a target word. Second, VWP studies tend to use four or more pictures, whilst IPL studies tend to use two pictures. A critical benefit of using more than two pictures is

that it permits simultaneous comparison of several types of experimental stimuli, while a distractor stimulus can still be included as a baseline measure. Automatic eye-tracking facilitates the identification of multiple fixation regions that are horizontally- and vertically-displaced from each other. Compared to their use with adults, the reliable use of automated eye-trackers with infants is a recent technological development. Typically, laboratories not using automated eye-trackers have relied on hand-coding of the eye-movements that are recorded by camera(s) located above or below the screen, permitting accurate identification of horizontally-displaced regions of interest but not vertically-displaced regions of interest.

In a standard IPL study, infants are shown a pair of pictures, e.g., dog and boat, and hear one of the pictures named in child-directed speech, e.g., "Look at the dog!". Provided they know the name of both pictures (as reported by their parents), infants spend more time fixating the named picture compared to the unnamed picture (Golinkoff et al., 1987), thereby indexing their comprehension of the name. IPL studies have also demonstrated that 14- to 24-month-old toddlers have phonetically-well specified representations of familiar words and are sensitive to mispronunciations (Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Swingley & Aslin, 2000).

However, few studies have examined the online mental processes of lexical access in toddlers. Do toddlers simply match speech sounds with concepts, or do they process spoken-words like adults, showing sensitivity to the phonological and semantic relationships between words and their referents? IPL studies have also shown that toddlers by the age of two, like adults, process speech incrementally. For example, Swingley, Pinto, and Fernald (1999) found that 24-month-olds were delayed at identifying a target when the target and distractor had the same onset consonant, e.g., dog and doll, but not when the target and distractor rhymed with

each other, e.g., ball and doll (as it does in American English). Similarly, Fernald, Swingley, and Pinto (2001) demonstrated that 18- and 21-month-olds could identify the target referent after hearing approximately the first two phonemes of the word. Furthermore, toddlers have shown an adult-like tendency to preferentially fixate a semantically-related, but non-target picture. For example, after hearing the word ‘strawberry’, 36-month-old toddlers would look at the picture of a cookie (which is taxonomically-related to strawberry) more than the picture of a chair (Johnson, McQueen & Huettig, 2011). This finding indicates that toddler word recognition can result not only in the activation of meaning representations for the given word but also related meanings.

Various adaptations of the toddler IPL task provide further support for the view that toddlers are sensitive to the phonological and semantic relationship between words by the age of two. Using a picture prime presented in silence prior to a standard IPL presentation, Mani and Plunkett (2010) found that target recognition was impaired (at chance) in 18-month-olds when they saw an unrelated prime picture (e.g., cake) before hearing the target word (e.g., ‘bird’), but not affected when they saw a phonologically-related prime picture (e.g., bib). This result could only have been obtained if toddlers had extracted the phonological codes of the prime picture and that the extracted phonological codes influenced processing of the heard target label. Other studies have investigated toddlers’ sensitivity to semantic relations between words using a similar procedure. These studies show that toddlers respond differentially when a target word is preceded by a semantically-related prime word compared to an unrelated prime (Arias-Trejo & Plunkett, 2009, 2013; Styles & Plunkett, 2009). For example, Arias-Trejo and Plunkett (2009) found that 21-month-olds (but not 18-month-olds) look longer at a target picture of a cat when they heard “Yesterday, I saw a dog (prime)! Cat (target)!” compared to “Yesterday, I saw a plate

(prime)! Cat (target)!". These findings reflect the formation of lexical-semantic links in the 21-month-olds' lexicon. Similar lexical priming studies show that target recognition is impaired when 24-month-olds hear an unrelated prime word, as opposed to a taxonomically- or thematically-related prime word (Arias-Trejo & Plunkett, 2013), indicating that 24-month-olds are sensitive to word pairings that are purely taxonomically-related (e.g., 'biscuit' and 'carrot') or purely thematically/associatively-related (e.g., 'rabbit' and 'carrot').

Mani, Durrant, and Floccia (2012) investigated the activation of phonological and semantic representations using the picture priming paradigm. Participants were presented with either a phono-semantic prime or an unrelated prime picture in silence, followed by the spoken-target word and a pair of target and distractor pictures. For example, they compared Boat (phono-semantic prime picture)–'Hat (target word)!' with Boat (unrelated prime picture)–'Table (target word)!', and found that target recognition in 24-month-olds was impaired (at chance) if participants saw the unrelated prime picture, but was not affected when participants saw the phono-semantic prime picture. The authors argue that these results demonstrate that, as with adults (see Marslen-Wilson & Zwitserlood, 1989), spoken-word recognition during infancy involves cascaded and co-activation of phonological and semantic representations (i.e., seeing a boat leads to the activation of coat which then pre-activates hat)³.

In sum, previous IPL studies suggest that, the toddler lexical-semantic system is initially different from that of adults. While toddlers demonstrate sensitivity to phonetic details and can match words with meanings by the age of one, they only begin to show adult-like appreciation for phonological and semantic similarities between words by the age of two, suggesting a

³ Note that since the target word was explicitly named, these results could be interpreted as the separate effects of phonological priming (i.e., seeing boat activated 'coat') and semantic priming (i.e., 'coat' did not inhibit the target word 'hat'), and may not necessarily be the result of cascaded activation of phonological and semantic representations.

qualitative shift from a toddler lexical-semantic system that encodes words in isolation to one that encodes words based on their phonological and semantic similarities. While spoken-word recognition in two-year-olds involves the automatic activation of phonologically- and semantically-related representations, the order and time course in which phonological and semantic representations are activated remains unclear. Are phonological and semantic representations activated at the same time or at different stages?

In principle, it is possible to examine the *relative* potency of phonological and semantic information using two-picture tasks. For example, a ‘target absent’ design might name an absent object ‘dog’ in the presence of pictures of a door and a cat. Any relative preference for one of the pictures over the other might indicate the degree to which phonological versus semantic/perceptual information drives looking behaviour. The relative timing of the eye fixations on the two pictures might also provide useful clues to the speed with which phonological and semantic information is activated. A tendency to fixate the door before the cat would point to earlier activation of phonological information compared to semantic information. However, in this two-picture IPL task, it is possible that toddlers show preferential fixation towards a phonological distractor, not because of the activation of phonological information, but just because the semantic/perceptual picture (cat) does not match the word ‘dog’ (the door is the only other picture on display). Inferring the activation of phonological information requires that the door is preferentially fixated *relative* to an *unrelated* picture, e.g., a plate. Since there are only two pictures on display, the standard IPL task does not permit simultaneous examination of the time course and competition between the activation of phonological and semantic representations during lexical access.

The Current Study

The two experiments presented here examine the ‘tug of war’ during spoken-word recognition between the phonological and semantic levels of representation. Experiment 1 investigates the time course of the activation of phonological and taxonomic representations in 24- and 30-month-olds. Experiment 2 examines the time course of the activation of phonological and thematic representations in a different sample of toddlers ranging between 25- to 30-months. Previous toddler studies have shown that both taxonomic and thematic relationships are involved in constituting a semantic network (Arias-Trejo & Plunkett, 2009, 2013; Styles & Plunkett, 2009). We, therefore, expect that both taxonomic and thematic representations act as potential competitors during the tug of war of spoken-word recognition. The design of both experiments was based on the target-absent design in Experiment 1 of Huettig and McQueen (2007)⁴.

The current study has two aims. The main aim is to provide direct evidence of the time course of the activation of phonological and semantic representations during lexical access in two-year-olds by adapting the adult four-picture VWP (Huettig & McQueen, 2007) to toddlers. No previous studies have demonstrated the utility of the four-picture VWP in measuring language-mediated visual attention in toddlers. The advantage of developing a toddler adaptation of the four-picture VWP is that it allows the incorporation of *unrelated* distractors as a baseline in addition to phonological and semantic distractors, thereby permitting the investigation of these competing types of representations in a way not possible in a standard two-picture IPL task.

⁴ We conducted a target-present experiment as a ‘proof of concept’, since four-picture VWP tasks have not previously been used with toddlers. The results are presented in Appendix D. In a test trial, 24- and 30-month-old participants heard a target word, and saw a four-picture display including a target, phonological distractor, taxonomic distractor and unrelated distractor. We successfully demonstrated that a four-picture target-present display can capture toddlers’ preferences for named target pictures. Both 24- and 30-month-old participants demonstrated a strong preference for the target over the unrelated distractor in response to the auditory label. There is tentative evidence that participants treated the related and unrelated distractors differently. However, it is also evident that the majority of participants’ attention was drawn to the target upon hearing the target label, leaving little room for the phonological and taxonomic distractors to play any substantial role in modulating toddlers’ visual preferences. It is therefore important to use a target-absent four-picture display to maximise toddlers’ fixations towards the phonological and taxonomic distractors.

Using a task similar to that used on adults enables us to investigate directly whether toddlers show an adult-like sensitivity to phonological and semantic relations between spoken-words and their referents. In particular, studying phonological versus semantic competition can inform us whether spoken-word recognition at this stage of development involves cascaded activation of phonological and semantic knowledge, and the time course and order of these activations, rather than a simple association between a label and its referent.

Based on Huettig & McQueen's (2007) findings, we predicted that both the 24- and 30-month-old participants would fixate the phonological distractor more than the unrelated distractor early in the time course—an early phonological preference, and the semantic distractor more than the unrelated distractor later in the time course—a late semantic preference. The finding of a phonological preference and/or a semantic preference would indicate the activation of both phonologically- and/or semantically-related representations during lexical access. In contrast, any lack of preference for either target-related object(s) would suggest lack of activation of phonologically- and/or semantically-related representations. Moreover, the order of phonological and semantic preferences would indicate the order in which different types of representations are activated during lexical access. An early phonological preference and a late semantic preference would indicate the cascaded activation and matching of phonological information followed by semantic information. Any lack of differences in the time course of a phonological and semantic preference would suggest that phonological and semantic knowledge are activated and matched in parallel.

The secondary aim is to explore the relationship between toddlers' age and vocabulary size, and language-mediated visual attention. How does language mediate visual attention for the toddler and how does the ability to extract phonologically- and semantically-related information

from speech influence the way toddlers direct their attention to objects in the world? A number of studies have reported that toddlers with a larger vocabulary size are faster at recognising spoken words in preferential looking studies (e.g., Fernald, Perfors, & Marchman, 2006; Fernald et al., 2001; Zangl, Klarman, Thal, Fernald, & Bates, 2005). Moreover, studies have shown that the N400-like semantic priming effect is stronger in toddlers with a larger vocabulary (Friedrich & Friederici, 2004; Rama, Sirri & Serres, 2013). It is likely that toddlers with a greater vocabulary have more experience with hearing, seeing and speaking, thereby strengthening the lexical representations of words and concepts in the mental lexicon. Such strengthened representations may allow faster retrieval and activation of related representations during spoken-word recognition. In the current study, we chose to study 24- and 30-month-old participants to identify any developmental trends across an age range when vocabulary is expanding at a rapid rate. Both age groups have relatively large receptive vocabularies (Mayor & Plunkett, 2011). Note that four-picture studies require double the number of stimuli compared to two-picture studies. While 30-month-olds have a substantially larger vocabulary size than 24-month-olds, we only included trials in which toddlers were familiar with all the words and pictures used, as reported by their parents (Hamilton, Plunkett, & Schafer, 2000). Therefore, any behavioural differences between the two age groups cannot be attributed to the lack of knowledge of individual words *per se*. Instead, any behavioural differences between the two age groups might reflect quantitative and/or qualitative differences of the lexicon, such as overall vocabulary size, robustness of lexical representations and the emergence of lexical-semantic structure. Based on previous toddler findings, we predicted that toddlers with greater language experience (either reflected by age or vocabulary size) would demonstrate greater preference for the target-related distractors over the unrelated distractors.

Statistical Methods

It is commonplace in infant IPL studies to analyse eye-gaze data by aggregating total fixation times at the target (T) and distractor (D) to obtain a proportional measure of target picture fixation (P) after target-word onset as shown in Equation 1.

$$P_{Target} = T / (T + D) \quad (1)$$

Since there are just two pictures to fixate, significant deviations from chance (0.5) are taken as evidence of a systematic response to the target word⁵. This method, however, ignores the *time course* of fixations during a trial that can be highly informative. For example, a total proportion of 50% of target fixation might result from an increase from 0 to 100% target looking, or a decrease from 100% to 0% target looking during the observed time window. Such timing differences have the potential to reveal the manner in which lexical representations are activated.

In contrast, adult VWP studies often aim to take into account more dynamic changes in eye fixations after target word onset. Post-naming fixations, usually 1000-2000ms after target word onset, are usually divided into finer timebins of equal duration, e.g., 50ms or 100ms each. In early studies (e.g., Allopenna et al., 1998; Huettig & McQueen, 2007), participants' preference for a target-related distractor over an unrelated distractor in each timebin was summarised by a 'preference ratio'. The preference ratio in each timebin can be compared to chance (0.5) in a one-sample t-test. A ratio significantly greater than 0.5 indicates a stronger preference for the target-related distractor over the unrelated distractor in a particular timebin. Growth of fixation preference is indexed by an increase in the number of significant t-tests

⁵ Note that, in principle, *any* significant deviation from chance—positive or negative—constitutes a systematic response to the target word. The typical finding is that toddlers show a preference for the target picture. Some investigators also take account of looking prior to the onset of the target word in order to obtain a baseline preference for each picture. Other studies use difference measures between single longest looks at the two pictures or latency to disengage from the target or distractor pictures.

across all timebins. While Bonferroni corrections can be applied to counteract the problem of multiple t-test comparisons, this approach is often too conservative for studies with strong hypotheses. Another limitation with this method of analysis is that eye-tracking data, like other response data, is binomial: a participant either looked (i.e., 100% looking) or did not look (i.e., 0% looking) in each frame recorded by the eye-tracker (for a 120Hz sampling rate, duration of one time frame is 8.3ms). As a result, eye-tracking data is unlikely to be normally distributed, especially in small timebins. Moreover, t-tests do not allow inspection of interactions, such as the effect of Condition and Age, and they do not account for individual variability among participants.

To address these issues, recent adult VWP and infant IPL studies have begun to analyse eye-tracking data using generalised-linear mixed models (e.g., Barr, 2008; Law II & Edwards, 2014; Von Holzen & Mani, 2012). In the current study, we used logistic mixed effects models with the R package glmmPQL, with penalised quasi-likelihood estimation. As the data follows a curvy trend with two peaks, we modelled time-course data using third-order (cubic) orthogonal polynomials (see chapter 3.3.3 in Mirman, 2014 for the advantages of using orthogonal polynomials over natural polynomials). Note that due to centering, the orthogonalised intercept term corresponds to the overall average, rather than the baseline of the measurement. The model has two levels⁶: Level 1 examines how individuals' fixation patterns change over time (linear), time² (quadratic) and time³ (cubic). Level 2 examines the effects of picture role, age group or vocabulary size on individual patterns of fixation. As an increased amount of attention to one object will inevitably lead to a decreased amount of attention to the other object(s) within the same timebin, we account for such dependence by including picture co-occurrence in each

⁶ See Equations 1-5 in Appendix A for the level-1 model, and level-2 model for the intercept, linear, quadratic and cubic terms respectively.

timebin nested under participant as random effects in the model.

The raw fixation data is initially organised by time frame⁷. The probability (p) of a participant fixating an area of interest (AOI) such as the target picture during a particular timebin, say 100ms, is determined by counting the number of frames in which the participant looked at the AOI (S) divided by the total number of frames recorded in the timebin (N), i.e., $p = S/N$. The probability of a participant not fixating the AOI (q) is then just $q = 1 - p$. Binomial logistic models estimate the log of the odds ratio for fixating and not fixating the AOI, i.e., $\log(p/q)$. It is important to note that eye-tracking gaze data, like other types of time-series data, can be autocorrelated, e.g., gaze at time point 2 may be dependent on gaze at time point 1. To remove the autocorrelation⁸ between the raw gaze data points, the data is aggregated in each timebin for each picture role and participant. To illustrate, assume two trials are recorded from a participant. Each trial lasts 1200ms and is divided into twelve 100ms timebins for growth curve analysis. The number of frames recorded (N) in each 100ms timebin in each trial would be $100\text{ms}/8.3\text{ms} = 12$ frames. The aggregated N would then be $2 \text{ trials} \times 12 \text{ frames} = 24$ frames. Suppose the participant fixated an AOI, such as the target picture, for 5 frames in timebin 1 during the first trial, and 10 frames in timebin 1 during the second trial. The aggregated S for timebin 1 would be $5 \text{ frames} + 10 \text{ frames} = 15$ frames. The probability of fixating the AOI in timebin 1 is $p = 15/24 = 0.625$ and the log odds ratio $\log(p/q) = \log(0.625/0.375) = \log(1.67)$.

Experiment 1

Experiment 1 compares the time course of phonological and taxonomic activation in 24- and 30-month-old toddlers. As illustrated in Figure 1, in a test trial, toddlers hear a target word

⁷ For example, within 100ms, an eye-tracker sampling at a rate of 120Hz records 12 frames of gaze data (8.3ms each).

⁸ Autocorrelation complicates analyses by reducing the number of independent observations and can also complicate the identification of significant covariance between time series.

(e.g., ‘bee’), and see a four-picture display containing a phonological distractor (e.g., bus), taxonomic distractor (e.g., cat), and two unrelated distractors (e.g., dress and sandwich). To ensure any preference for the taxonomic distractor we observe is due to toddlers’ appreciation of the taxonomic relationship between the target word and the taxonomic distractor, rather than by association, we checked that all the taxonomic distractors were unassociated based on adult association norms (Kiss, Armstrong, & Milroy, 1972; Nelson, McEvoy, & Schreiber, 1998).

Figure 1. Example of the four-picture array. The target is ‘bee’, the phonological distractor is bus, the taxonomic distractor is cat, and the unrelated distractors are sandwich and dress.

Method

Participants. Twenty-four 24-month-olds ($M_{age} = 24.24$, age range = 23.28–25.44, Male = 15, Female = 9) and twenty-four 30-month-olds ($M_{age} = 29.88$, age range = 29.4–30.96, Male = 10, Female = 14) recruited from British English monolingual households were included in the analysis. Nine additional participants were excluded from analysis due to fussiness. A minimum of eight test trials (i.e. one-third of test trials) is required from each participant for analysis. Data from six additional participants was excluded from analysis due to unfamiliarity (receptive vocabulary) with the majority of the words used in the current experiment (as reported by the parents of the toddler participants), which resulted in less than eight test trials available for analysis.

Design. Each participant saw a total of 24 target-absent test trials and 13 target-present filler trials, which were presented in a pseudo-random order in four blocks. The filler trials served to

maintain the participants' attention, as toddlers at this age can easily get frustrated if the target word never matches any of the pictures⁹. Participants were randomly assigned to one of four lists (see Tables A2 to A5 in Appendix A). The four lists contained the same trials but served to counterbalance for block (yoke set) order and picture location. Each list started with a filler trial, which was the same trial for all four lists. In each block, there was always a filler trial before the first test trial and after the second test trial. As shown in Figure 2, there was an attention getter at the beginning of each trial. The four pictures appeared on the screen for a total of 8000ms. The target word onset, e.g., 'bee', was always at 4000ms, and was preceded by a prompt phrase "Look at the". Target word onset thus divided the trial into two phases, a 4000ms pre-naming phase and 4000ms post-naming phase.

Figure 2. Experiment 1 trial timeline. Each trial began with a 1000ms attention getter. The visual stimuli appeared for a total of 8000ms. Onset of the auditory label of the target word was always at 4000ms.

Stimuli. The test stimuli were six sets of four pictures (total 24 pictures), and 24 target words (presented as an auditory label). As shown in Table A1 in Appendix A, each of the six sets of

⁹ While the four-picture paradigm allows us to make two or more comparisons simultaneously, it also poses restrictions on the experimental design. For example, toddler vocabulary is limited and the use of four pictures means that more words had to be used than in a two-picture study. Moreover, toddlers can get upset if there is no referent to the spoken-word in a target-absent set-up. However, we checked that the toddler participants of this study fixated the target during the target-present filler trial, suggesting that they continued to search for the target picture during the whole experiment. Researchers, therefore, have to balance between having enough target-present filler trials and target-absent test trials, and succeeding in getting toddlers to complete the experiment.

pictures were yoked¹⁰ four times, and each time they were paired with a different target word. As a result, each picture in each set served four picture roles: once as a phonological distractor to the target word, once as a taxonomic distractor to the target word, and twice as an unrelated distractor. The phonological distractor depicts a picture that, if labelled, starts with the same onset consonant as the target word. The taxonomic distractor is in the same taxonomic category as the target. As there are no existing category norms for toddlers, the taxonomic distractors were chosen based on previous toddler studies that have demonstrated semantic priming effects with these items (e.g., Arias-Trejo & Plunkett, 2013), and category norms for verbal items in Grade 2-6 children (Posnansky, 1978). The unrelated distractors were unrelated to each other, and unrelated to the target, phonological distractor and taxonomic distractor. Pictures in each set (e.g., bus, cat, sandwich and dress) were unassociated according to the University of South Florida Free Association Norms (Nelson et al., 1998) and the Edinburgh Associative Thesaurus (Kiss et al., 1972), so that we can be confident that potential phonological and taxonomic effects are not driven by association.

Among the 48 pictures and target words (ranging from one to four syllables), 40 were chosen from the OCDI (Hamilton et al., 2000), while eight of them were from extra word norm data that we have collected alongside the OCDI since 2012. Target words and pictures in the test trials were chosen based on the criterion that at least 60% of 24-month-olds were reported to understand the word. There were in total eight categories: animals, clothing, electrical appliances, food, furniture, toiletries, toys, and vehicles.

Visual Stimuli. Realistic photographic representations were used. Objects were edited out of

¹⁰ The yoked control procedure was first used in operant conditioning experiments with animals. In a typical operant conditioning experiment, the same reinforcement (e.g., food) is ‘yoked’ (paired up) with either the experimental contingency (e.g., food delivery is contingent on lever pressing) or control contingency (e.g., food delivery is not contingent on lever pressing). In a typical two-picture infant IPL study, yoking typically refers to the counterbalancing of the role of the picture stimulus (e.g., once as target, once as distractor).

their original background, and placed in the centre of a 9.53 x 9.53cm grey background using Adobe Photoshop (50% grey colour). The four square pictures (i.e., the four AOIs of eye-gaze measurement), were located in the upper left, upper right, lower left and lower right corners of an invisible square, of a total area of 28.58 x 28.58cm (23.7°), with a distance of 9.53cm (8.34°) separating the pictures horizontally and vertically. See Appendix E for adult visual similarity ratings of the target referent and each type of distractor pictures.

In each set of four pictures, each object had a different main colour. Objects without a typical colour were assigned a colour that had not been used within the same set, e.g., a car does not have a typical colour. However, objects with a typical colour(s) were always represented in their typical colour(s), e.g., a banana is yellow. The heads of all animals and the front of all vehicles faced towards the middle to eliminate the possibility of picture orientation or head direction as a cue.

Auditory Stimuli. A Southern British English speaker recorded auditory labels of target words preceded by a carrier phrase ‘Look at the’ in child-directed speech. Mean duration of target words was 602ms. Background noises, and head and tail clicks, were removed using GoldWave.

Picture Location. Picture location was counterbalanced both within and between participants. Within participants, each picture (e.g., car) appeared at each of the four screen locations once across the four presentations. Between participants, pictures in each trial appeared at all four locations for an equal number of times. Location of the pictures by picture role (i.e., the phonological distractor, taxonomic distractor, unrelated distractor 1 and unrelated distractor 2) was also counterbalanced both within and between participants. Within participants, each picture role appeared at either the left or right side of the screen an equal number of times (i.e., 12 times on the left and 12 times on the right). Between participants, the picture roles in each trial were

counterbalanced, and appeared at all four locations on the screen an equal number of times.

There were at least 2 test trials and a filler trial between repetitions of any picture set.

Procedure. Toddlers sat on their caregiver's lap approximately 65cm from a Tobii TX 300 eye tracker and a 23-inch screen of 1920 x 1080 screen resolution. The sampling rate was 120Hz.

The study was run using PresentMate, a custom Matlab stimuli presentation framework based on Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

Communication with the Tobii eye-tracker was implemented using the Tobii Analytics Matlab binding based on Tobii Analytics SDK 3.0. A nine-point calibration was performed. The

calibration attention getter was a colourful beach ball. The background colour of the screen remained grey (same colour as the background of the visual stimuli) throughout the experiment.

Auditory stimuli were presented through a centrally-located loudspeaker. In an adjacent room, the experimenter monitored the toddler's behaviour through a centrally-located video camera

above the screen. To achieve good calibration, individual locations were recalibrated when

necessary. After calibration, toddlers were shown in total 37 trials (24 test trials). The

experimenter initiated all trials by pressing a computer key, when the toddler's attention was on the screen. Caregivers were instructed to keep their eyes shut, remain quiet and refrain from any

interactions with toddlers during the experiment. Eye-tracking data was also processed in Matlab

using custom code to extract fixation information. Minimum fixation duration was set to 100ms

and a radius of 35 pixels.

Results

Overall, 24-month-olds contributed 447 out of 576 trials (on average 19 out of 24 trials per participant) for analysis. One-hundred and ten trials (19%) were excluded because of unfamiliarity with pictures or words used. Nine trials (2%) were excluded due to zero post-

naming looking time. A further 10 trials (2%) did not provide any data due to the toddler not looking during the entire trial. Thirty-month-olds contributed 513 out of 576 trials (on average 21 out of 24 trials per participant) for analysis. Forty trials (7%) were excluded due to unfamiliarity with pictures or words used. Nine trials (2%) were excluded due to zero post-naming looking time. A further 14 trials (2%) did not provide any data due to the toddler not looking during the entire trial. The same pattern of results were obtained irrespective of the inclusion of words unknown to the toddler. However, only results for the words known to the toddler are included as these provide the clearest findings.

Figure 3a shows the grand mean of the time course of fixations for the phonological distractor, taxonomic distractor and an average of the two unrelated distractors, aggregating participants from both age groups. The figure indicates that participants fixated the phonological distractor more than the unrelated distractors earlier in the time course, and fixated the taxonomic distractor more than the unrelated distractors later in the time course.

Figure 3. Grand mean of time course of fixation from onset of the auditory label of the target in (a) Experiment 1 and (b) Experiment 2.

Growth Curve Analysis.

We determined the starting point for analysis by locating the breakpoint at which one of the curves showed apparent change. As seen in Figure 3a, the earliest break was at around 330ms after target word onset. Fixations for the four objects became approximately asymptotic from 1500ms to the end of the trial. A time window of 1200ms from 330 to 1530ms after target word

onset thus captures both age groups' fixation patterns. The time course data was first modelled using third-order orthogonal polynomials with fixed effects of picture role (phonological vs. taxonomic vs. unrelated) on all time terms, age group (24- vs. 30-months) on the intercept and linear terms, and picture co-occurrence in each timebin nested under participant random effects on the intercept and slope terms. The unrelated distractor and the 30-month-old age group were treated as the baseline in the model, and relative parameters were estimated for the phonological and taxonomic distractors, and the 24-month-old age group.

Figure 4 shows the data as points and the model as lines. The data points indicate that participants in both age groups fixated the phonological distractor more than the unrelated distractors earlier in the time course, and fixated the taxonomic distractor more than the unrelated distractors later in the time course. Table 1 summarises the results of the statistical model. For the phonological distractor, the significant intercept term shows that participants spent more time fixating the phonological distractor than the unrelated distractor. The significant linear, quadratic and cubic terms indicate that fixations towards the phonological and unrelated distractors followed different trajectories: phonological distractor fixation peaked shortly after target-word onset, while fixations towards the unrelated distractor remained relatively flat. There was no reliable phonological distractor x age interaction (Phonological–Unrelated*24-months – Phonological–Unrelated*30-months).

Figure 4. Experiment 1: Growth curve analysis of (a) 24-month-old and (b) 30-month-old participants. Data shown in points, the model shown in lines.

Table 1

Experiment 1: Growth Curve Analysis by Age Group.

Fixed Effects	Model Terms	Estimates	SE	z	p
Phonological–Unrelated	Intercept	0.15	0.04	4.04	< .001***
	Linear	-0.44	0.13	-3.39	< .001***
	Quadratic	-0.25	0.10	-2.66	.007**
	Cubic	0.33	0.10	3.52	< .001***
Taxonomic–Unrelated	Intercept	0.23	0.04	6.25	< .001***
	Linear	0.78	0.13	6.16	< .001***
	Quadratic	-0.03	0.09	-0.28	.781
	Cubic	-0.08	0.09	-0.83	.411
Taxonomic–Unrelated – Phonological–Unrelated	Intercept	0.08	0.04	2.22	.027*
	Linear	1.22	0.13	9.73	< .001***
	Quadratic	0.23	0.09	2.48	.013*
	Cubic	-0.41	0.09	-4.49	< .001***
24-months–30-months	Intercept	-0.00	0.05	-0.05	.962
	Linear	-0.10	0.14	-0.72	.473
Phonological–Unrelated *24-months – Phonological–Unrelated *30-months	Intercept	-0.05	0.05	-0.86	.388
	Linear	0.22	0.19	1.16	.247
Taxonomic–Unrelated *24-months – Taxonomic–Unrelated *30-months	Intercept	0.05	0.05	0.89	.374
	Linear	-0.02	0.19	-0.10	.922

*** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$

There was also a systematic preference to fixate the taxonomic distractor. The significant intercept term shows that participants spent more time fixating the taxonomic distractor than the unrelated distractor. The significant linear term confirms that fixations towards the taxonomic distractor gradually increased during the time course while the slope of the unrelated distractor remained relatively flat. There was no reliable taxonomic distractor x age interaction (Taxonomic–Unrelated*24-months – Taxonomic–Unrelated*30-months).

To compare the time course of phonological and taxonomic fixations, we computed the difference between the Phonological–Unrelated and Taxonomic–Unrelated comparisons (See Table 1, Taxonomic–Unrelated – Phonological–Unrelated). The significant intercept term shows that, overall, toddlers fixated the taxonomic distractor more than the phonological distractor. More importantly, phonological fixation peaked before taxonomic fixation. The significant linear, quadratic and cubic terms confirm that the growth trend of taxonomic fixation was significantly different from that of phonological fixation. As seen in Figure 5, if we remove the significant linear, quadratic and cubic terms for *picture role*, the reduced model (in grey) would predict no differences between the time course of phonological and taxonomic fixations.

Figure 5 Experiment 1: Different time course of fixation for the phonological and taxonomic distractors in (a) 24-month-old and (b) 30-month-old participants. Data shown in points, the full model shown in coloured lines, the reduced model shown in grey lines.

Vocabulary Analysis. For the twenty-four 24-month-olds included in the analysis, the averaged OCDI receptive vocabulary score was 439 out of 553 words ($SD = 46$), and the averaged OCDI expressive vocabulary score was 276 out of 553 ($SD = 130$). For the twenty-four 30-month-olds included in the analysis, the averaged CDI receptive vocabulary score was 491 out of 553 ($SD = 33$), and the averaged OCDI expressive vocabulary score was 457 out of 553 ($SD = 70$). Twenty-four-month-olds had significantly lower OCDI receptive ($U = 112.5, p < .001$) and expressive vocabulary scores ($U = 93, p < .001$) than 30-month-olds. The Spearman's rho signed-rank test revealed a significant correlation between participants' receptive and expressive OCDI scores ($r_s = .78, N = 48, p < .001$).

We examined whether measures of vocabulary size were predictive of phonological and taxonomic fixation times. We first replaced age group with median-centred receptive vocabulary size as a continuous predictor in the model, and we subtracted the median receptive OCDI score

(469 words) of all 48 participants from each participant's OCDI receptive score¹¹. The data was modelled using third-order orthogonal polynomials with fixed effects of picture role (phonological vs. taxonomic vs. unrelated) on all time terms, receptive vocabulary size on the intercept and linear time terms, and picture co-occurrence in each timebin nested under participant as random effects on the intercept and linear terms. The unrelated distractor and participants with a smaller vocabulary were treated as the baseline in the model, and relative parameters were estimated for the phonological and taxonomic distractors, and participants with a larger vocabulary.

Table 2 summarises the results for the fitted model. As in the previous model, there were significant effects of picture role. Regardless of receptive vocabulary size, participants fixated the phonological and taxonomic distractors more than the unrelated distractors. There was a significant phonological fixations x receptive vocabulary interaction (Phonological–Unrelated *Larger Vocabulary – Phonological–Unrelated *Smaller Vocabulary, see also Figure 6a). The significant intercept term indicates that participants with a larger receptive vocabulary size overall spent more time fixating the phonological distractor than those with a smaller receptive vocabulary. Moreover, the significant linear term indicates that the negative slope of phonological fixation was substantially steeper for participants with a larger receptive vocabulary than those with a smaller receptive vocabulary. There was also a significant taxonomic fixation x receptive vocabulary interaction (Taxonomic–Unrelated *Larger Vocabulary – Taxonomic–Unrelated *Smaller Vocabulary, see also Figure 6b). The significant intercept term indicates that participants with a larger receptive vocabulary overall spent less time fixating the taxonomic distractor than those with a smaller receptive vocabulary.

¹¹ For ease of presentation, we divided the median-centred score by 100 to reduce the number of decimal places of the model's estimated values, e.g., 0.0005457 to 0.05457. This does not change the *Z* and *p* values of the predictions.

Table 2

Experiment 1: Growth Curve Analysis by Receptive Vocabulary Size.

Fixed Effects	Model Terms	Estimates	SE	z	p
Phonological–Unrelated	Intercept	0.12	0.03	4.48	< .001***
	Linear	-0.31	0.09	-3.29	.001**
	Quadratic	-0.25	0.09	-2.72	.006**
	Cubic	0.33	0.09	3.58	< .001***
Taxonomic–Unrelated	Intercept	0.26	0.03	9.65	< .001***
	Linear	0.78	0.09	8.48	< .001***
	Quadratic	-0.03	0.09	-0.31	.762
	Cubic	-0.08	0.09	-0.84	.402
Taxonomic–Unrelated – Phonological–Unrelated	Intercept	0.14	0.03	5.18	< .001***
	Linear	1.09	0.09	11.94	< .001***
	Quadratic	0.23	0.09	2.51	.012*
	Cubic	-0.41	0.09	-4.56	< .001***
Larger Vocabulary (Receptive) – Smaller Vocabulary (Receptive)	Intercept	0.05	0.06	0.94	.348
	Linear	0.17	0.16	1.06	.290
Phonological–Unrelated *Larger Vocabulary – Phonological–Unrelated *Smaller Vocabulary	Intercept	0.16	0.07	2.51	.012*
	Linear	-0.61	0.23	-2.70	.007**
Taxonomic–Unrelated *Larger Vocabulary – Taxonomic–Unrelated *Smaller Vocabulary	Intercept	-0.25	0.06	-3.95	< .001***
	Linear	0.23	0.22	1.09	.276

*** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$

Figure 6. Experiment 1: Growth curve analysis with receptive vocabulary size as continuous predictor. (a) Phonological distractor fixation and (b) Taxonomic distractor fixation.

We also attempted to fit the model using expressive vocabulary size as a continuous

predictor. However, despite showing the same data trend, there were no significant interactions between expressive vocabulary size and phonological or taxonomic fixations.

Discussion

Experiment 1 demonstrated that there is generalised phonological and taxonomic activation during lexical access. Both 24- and 30-month-olds preferentially fixated the phonological distractor immediately after target-word onset (330ms to 1130ms), suggesting that the target word activates phonologically-related words, e.g., ‘bee’ activates ‘bus’. Our toddler participants also preferentially fixated the taxonomic distractor, though somewhat later in the time course of the trial (630ms onwards), reflecting the activation of the taxonomic category associated with the target word and the taxonomic distractor, e.g., ‘bee’ activates the taxonomic category *animal* leading to the activation of cat. It is also possible that the activation occurs at the lexical level, rather than the conceptual level. For example, fixations may be driven by the lexical-semantic links between ‘bee’ and ‘cat’ in the mental lexicon. It is noteworthy that our toddler participants demonstrated an early preference for the phonological distractor over the taxonomic distractor similar to that observed for the adults in Huettig and McQueen’s (2007) study. This finding indicates that toddlers, like adults, extract phonological information from the spoken word faster than they extract semantic information.

Vocabulary Size. We found that vocabulary size, rather than age, was a better predictor of individual differences in this experiment. Participants with a larger receptive vocabulary displayed a stronger and longer phonological preference than participants with a smaller receptive vocabulary (see the interaction on the intercept and linear terms in Table 2). The larger vocabulary participants devoted more attention to the phonological distractors early in the trial. Consequently, they began to attend the taxonomic distractor later in the trial (see the interaction

on the intercept term in Table 2). There is accumulating evidence that age-related vocabulary growth is related to efficiency in spoken-word recognition during early childhood. For example, faster response speeds and greater target-looking time during IPL spoken-word-recognition tasks as vocabulary size increases have been reported during and beyond the second year (Fernald et al., 2006; Fernald et al., 2001; Zangl et al., 2005, Law II & Edwards, 2014). Moreover, some studies investigating novel word learning report that 14-month-olds' novel word learning performance is correlated with receptive and expressive vocabulary size (Werker, Fennell, Corcoran, & Stager, 2002).

In the current study, toddlers with a larger vocabulary were likely to have known the test words for a longer time, and had more experience hearing and saying these words, which in turn may have strengthened the phonological representations and word connections in the mental lexicon. Consolidation of the phonological representation of words may in turn lead to easier and more rapid retrieval of names in response to viewing a picture, or retrieving other lexical candidates when hearing the auditory label, and hence faster and longer responses to the phonological distractor. Larger vocabularies may also lead to a more *refined* phonological representation of words as denser lexicons require more detail to discriminate between similar words (cf. Metsala and Walley's (1998) Lexical Restructuring Model). Detailed phonological representation of words can in turn lead to enhanced early activation of alternative lexical candidates and hence a stronger pull towards phonological competitors¹².

Experiment 2

In Experiment 2, we ask whether another type of semantic information, *thematic/associative*, is extracted during word recognition such that a thematic distractor can

¹² Of course, another consequence of detailed phonological representations would be faster rejection of lexical competitors once sufficient auditory information has been processed.

attract fixations in the same manner as the taxonomic distractor in the previous experiment. The design of Experiment 2 is largely similar to that of Experiment 1, except that we replaced the taxonomic distractor with a thematic distractor. The thematic distractor was defined as an object or concrete context associated to the target (e.g., hair and brush) according to the University of South Florida word association norms (Nelson et al., 1998) and the Edinburgh Associative Thesaurus (Kiss et al., 1972), and what the authors deemed as thematically-related. Thematically-related objects are not necessarily associated (e.g., spoon and dessert). However, since thematic relations are largely based on personal experiences and may vary greatly among individuals, word association is a useful indicator of the typicality of the thematic relations. Importantly, the objects chosen as thematically-related distractors are taxonomically-unrelated according to adult standards.

We predicted that toddlers would preferentially fixate the phonological distractor early in the time course of a trial. As previous studies have established that toddlers by the age of two possess and utilise knowledge of thematic relations (e.g., Scott, Serchuk, & Mundy, 1982), we also predicted that toddlers in Experiment 2 would preferentially fixate the thematic distractor over the unrelated distractor. Moreover, based on Huettig and McQueen (2007), we expected to observe a late thematic preference. As Experiment 1 showed that vocabulary size was a better predictor than age in predicting the dynamics of phonological and semantic activation, we relaxed the age criterion to between 25- to 30-months in an attempt to cover a larger range of participant vocabulary size. Finally, a comparison of Experiments 1 and 2 would enable us to examine the relative time course and strengths of taxonomic and thematic activation in response to a comprehended word.

Method

Participants. Twenty-four participants between 25- and 30-months-old ($M_{age} = 26.88$, age range = 25.44–30.48, Male = 12, Female = 12) were recruited from British English monolingual households. Three participants were excluded due to fidgeting during testing. One participant was excluded as a result of not knowing most of the words used in this experiment, leaving only one trial available for analysis.

Design. The design of Experiment 2 was identical to Experiment 1, except that the post-naming phase lasted for 3000ms instead of 4000ms. Note that all of the effects reported in the previous experiments took place within 2000ms after word onset. See Tables B2 to B5 in Appendix B for the stimuli lists.

Stimuli. Among the 24 pictures and 24 target words (ranging from one to four syllables), 39 were chosen from the OCDI (Hamilton et al., 2000), while nine of them were from extra word norm data that we have collected alongside the OCDI since 2012. Target pictures and words in the test trials were chosen based on the criterion that at least 60% of 24-month-olds were reported to understand the word. Each trial included a phonological distractor (e.g., house), which shared the same onset consonant as the target word (e.g., hair); a thematic distractor (e.g., brush), which belonged to the same thematic category, and was associated to the target according to the Edinburgh Associative Thesaurus (Kiss et al., 1972) and the University of South Florida Free Association Norms (Nelson et al., 1998); and two unrelated distractors (e.g., flower and camera), which were unrelated to each other, and to the target, phonological distractor and thematic distractor. The phonological and unrelated distractors were unassociated to the target. See Table B1 for the yoked stimuli list.

Auditory Stimuli. The auditory stimuli were produced in the same way as in Experiment 1. In each trial, participants heard a sentence that included a prompt phrase followed by a target word,

e.g., ‘Look at the bath!’ . Mean duration of target words was 797ms.

Procedure. The procedure was identical to Experiment 1.

Results

There were in total 407 out of 576 trials available for analysis (on average 17 out of 24 trials per participant). A total of 104 trials (18%) were excluded due to unfamiliarity with the pictures or words used as reported by parents. A further 65 trials (11%) did not record any data due to toddler fidgeting, or were excluded due to zero looking time or bias for one picture throughout the entire trial. The same pattern of results were obtained irrespective of the inclusion of words unknown to the toddler. However, only results for the words known to the toddler are included as these give the clearest findings. Figure 3b shows the grand mean of fixation probabilities for each type of distractor and indicates that participants fixated the phonological distractor more than the unrelated distractors earlier in the time course, and fixated the thematic distractor more than the unrelated distractors later in the time course.

Vocabulary Analysis. The averaged OCDI receptive score was 436 out of 553 ($SD = 57$), and the averaged OCDI expressive score was 305 out of 553 ($SD = 120$). One participant was excluded from the vocabulary analysis because the parent did not complete the full OCDI, but only the experimental stimuli checklist during the visit. Participants’ OCDI receptive scores are correlated with their expressive scores ($r_s = .41$, $N = 23$, $p = .0506$).

Growth Curve Analysis

To facilitate comparison with Experiment 1, we analysed the data using the same time window: 330–1530ms after target onset. The overall time course of fixations was captured by third-order orthogonal polynomials with fixed effects of picture role (phonological vs. thematic vs. unrelated) on all time terms, either the median-centred receptive or expressive vocabulary

size as a continuous predictor on the intercept and linear terms, and picture co-occurrence in each timebin nested under participant as random effects on the intercept and linear terms. The unrelated distractor and participants with a smaller vocabulary were treated as the baseline in the model, and relative parameters were estimated for the phonological and taxonomic distractors, and the participants with a larger expressive vocabulary size. We found that only expressive vocabulary predicted both phonological and thematic fixation, but not receptive vocabulary.

Table 3 summarises the results of the statistical model. The significant phonological intercept term indicates that participants fixated the phonological distractor more than the unrelated distractor. There was a significant interaction between amount of looking at the phonological distractor and expressive vocabulary size (Phonological–Unrelated**Larger Vocabulary–Smaller Vocabulary*), indicating that participants with a larger expressive vocabulary overall spent more time fixating the phonological distractor than those with a smaller expressive vocabulary (see Figure 7a). Similarly, the significant thematic intercept term indicates that participants spent more time fixating the thematic distractor than the unrelated distractor. The significant linear term confirms that fixations towards the thematic and unrelated distractors followed different trends with fixations towards the thematic distractor increasing throughout the time window. Moreover, there was a significant interaction between amount of looking at the thematic distractor and expressive vocabulary size (Thematic–Unrelated**Larger Vocabulary–Smaller Vocabulary*). Overall, participants with a larger expressive vocabulary fixated the thematic distractor more than participants with a smaller expressive vocabulary (see Figure 7b).

Table 3

Experiment 2: Growth Curve Analysis by Expressive Vocabulary Size.

Fixed Effects	Model Terms	Estimates	SE	z	p
Phonological–Unrelated	Intercept	0.14	0.05	3.02	.003**
	Linear	-0.18	0.16	-1.08	.280
	Quadratic	0.19	0.16	1.19	.234
	Cubic	0.26	0.16	1.64	.101
Thematic–Unrelated	Intercept	0.47	0.04	10.37	< .001***
	Linear	1.51	0.16	9.60	< .001***
	Quadratic	0.28	0.15	1.82	.069
	Cubic	-0.04	0.15	-0.25	.806
Thematic–Unrelated – Phonological–Unrelated	Intercept	0.33	0.04	7.43	< .001***
	Linear	1.69	0.15	11.14	< .001***
	Quadratic	0.09	0.15	0.60	.546
	Cubic	-0.30	0.15	-2.00	.045*
Larger Vocabulary (expressive) – Smaller Vocabulary (expressive)	Intercept	-0.06	0.04	-1.61	.108
	Linear	-0.10	0.10	-0.99	.322
Phonological–Unrelated* Larger Vocabulary – Phonological–Unrelated * Smaller Vocabulary	Intercept	0.08	0.04	2.06	.040*
	Linear	-0.07	0.13	-0.53	.594
Thematic–Unrelated* Larger Vocabulary – Thematic–Unrelated* Smaller Vocabulary	Intercept	0.14	0.04	3.84	< .001***
	Linear	0.14	0.12	1.09	.275

*** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$

Figure 7. Experiment 2: Growth curve analysis with expressive vocabulary size as continuous predictor. (a) Phonological distractor fixation. (b) Thematic distractor fixation.

To compare the time course of phonological and thematic fixations, we computed the difference between the Phonological–Unrelated and Thematic–Unrelated comparisons (see Table 3, Thematic–Unrelated – Phonological–Unrelated). The significant intercept term indicates that,

overall, participants fixated the thematic distractor more than the phonological distractor. More importantly, as in Experiment 1, phonological fixation peaked earlier than thematic fixation. The significant linear term and cubic term confirm that the growth trend of thematic fixation was significantly different from that of phonological fixation.

Discussion

The toddlers in this experiment preferentially fixated the phonological distractor immediately after target word onset, reflecting early detection of a phonological match between the target word and the label associated with the phonological distractor picture, and replicating a similar effect reported in Experiment 1. As predicted, they also preferentially fixated the thematic distractor later in the trial, reflecting a match between thematic information associated with the target word and the thematic distractor, e.g., ‘hair’ activates the thematic information leading to fixation of the picture of brush. The ‘pull’ of the thematic distractor later in the trial resembles that of the taxonomic distractor in Experiment 1 suggesting that thematic/associative information is accessed later than phonological information when comprehending a word.

Experiment 2 also reflects the finding in Experiment 1 that early fixations of the phonological distractor were more pronounced for the larger vocabulary toddlers (though note that in this case, expressive vocabulary was the continuous variable entered into the model). We do not fully understand why phonological distractor fixations were predicted by receptive vocabularies in Experiment 1, but by expressive vocabularies in Experiment 2. Fixations towards the thematic distractor were also more pronounced for toddlers with a larger expressive vocabulary. One possibility is that Experiment 2 highlighted thematic/associative processing, and that productive word combinations permit the toddler to highlight the relationship between the items referenced in an utterance and so strengthen any associations between them. Increasing

expressive vocabulary size in toddlers commonly goes hand-in-hand with the emergence of combinatorial utterances (Bates, Bretherton, & Snyder, 1988). The positive relationship between expressive vocabulary size and amount of thematic fixation for the toddlers in this study might be an indirect result of their level of command of combinatorial language. In the absence of any direct evidence of these toddlers' combinatorial language, such explanations must, however, remain speculative.

Experiment 1 vs. Experiment 2. The design of Experiments 1 and 2 differ primarily in the status of the semantic distractor: taxonomic in the former and thematic and associative in the latter¹³. Since the design of the experiments was otherwise identical, it is meaningful to ask whether the two types of semantic distractors influenced eye-fixations in similar ways. It is apparent that both attracted the toddlers' attention approximately within 1000ms after target word onset, even though neither matched the target. We can also use growth curve analysis to ask whether taxonomic and thematic distractors are otherwise equivalent. We compared the findings of all participants in Experiments 1 and 2 with a mixed effects growth curve model using third-order orthogonal polynomials with fixed effects of picture role (phonological vs. semantic) on all time terms, experiment (Experiment 1 (taxonomic) vs. Experiment 2 (thematic)) on intercept and slope terms, and picture co-occurrence nested under participant as random effects on the intercept and linear terms. The modelled growth curves are depicted separately in Figure 8 for phonological distractors (left panel) and semantic distractors (right panel), each against their respective unrelated baseline distractors. As seen in Figure 8a, the fitted data suggest that participants' preference for the phonological distractor over the unrelated distractor followed a similar trend in both Experiments 1 and 2. In contrast, as seen in Figure 8b,

¹³ Different stimuli were used in Experiment 1 and 2, because it is difficult find stimuli in this age range which can act as both taxonomic and thematic distractors in a yoked fashion. For example, key can be a thematic distractor of door, but it is difficult to categorise key as a taxonomic distractor.

participants' preference for the thematic distractor in Experiment 2 increased at a faster rate and reached a higher peak than that for the taxonomic distractor in Experiment 1.

Figure 8. Experiment 1 vs. Experiment 2: Growth curve analysis of all participants in Experiment 1 and Experiment 2. Data shown in points, the model shown in lines for (a) Phonological distractor fixation against unrelated distractor fixation and (b) Taxonomic or thematic distractor fixation against unrelated distractor fixation.

Table 4 summarises the model's results. In both experiments, we observed an early phonological preference (significant intercept, linear and cubic terms) and a late semantic preference (significant intercept and linear terms). The non-significant effect of Experiment indicates that, overall, participants spent similar amounts of time looking at the four pictures in both experiments. The non-significant interaction term 'Phonological–Unrelated*Experiment 2 – Phonological–Unrelated*Experiment 1' confirms that participants showed similar levels of phonological preference in both experiments. However, the significant intercept term on the interaction 'Semantic–Unrelated*Experiment 2 – Semantic–Unrelated*Experiment1' indicates that there was an overall greater thematic preference than taxonomic preference. Moreover, the

significant linear term on the interaction indicates, as is readily seen in Figure 8b, that thematic fixations increased at a faster rate than taxonomic fixations.

Table 4

Experiment 1 vs. Experiment 2: Growth Curve Analysis.

Fixed Effects	Model Terms	<i>Estimates</i>	<i>SE</i>	<i>z</i>	<i>p</i>
Phonological–Unrelated	Intercept	0.13	0.03	3.98	< .001***
	Linear	-0.33	0.11	-2.94	.003**
	Quadratic	-0.05	0.08	-0.55	.581
	Cubic	0.29	0.08	3.43	< .001***
Semantic–Unrelated (Taxonomic and Thematic)	Intercept	0.25	0.03	7.92	< .001***
	Linear	0.77	0.11	7.00	< .001***
	Quadratic	0.13	0.08	1.58	.114
	Cubic	-0.06	0.08	-0.68	.496
Experiment 2–Experiment 1	Intercept	-0.06	0.05	-1.30	.195
	Linear	-0.23	0.13	-1.86	.063
Phonological–Unrelated*Exp. 2 – Phonological–Unrelated*Exp. 1	Intercept	-0.04	0.05	-0.84	.402
	Linear	0.17	0.17	1.02	.306
Semantic–Unrelated*Exp. 2 – Semantic–Unrelated*Exp. 1	Intercept	0.16	0.05	3.27	.001**
	Linear	0.72	0.16	4.40	< .001***
*** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$					

Early developmental research proposed that young children initially possess only thematic knowledge (Annett, 1959; Inhelder & Piaget, 1964), or that there is a developmental shift from a thematic preference to taxonomic preference from childhood to adulthood (Denney, 1974; Scott, Serchuk & Mundy, 1982; Smiley & Brown, 1979). However, later studies provided evidence against these views (e.g., Greenfield & Scott, 1986; Waxman & Namy, 1997). For example, Waxman and Namy (1997) demonstrated in a forced-choice task that 3- and 4-year-olds were more likely to make a taxonomic or thematic choice depending on how the task instructions were phrased. Two-year-olds, in contrast, showed no systematic preference for either type of relations regardless of task instructions.

Our results show that the activation of taxonomic and thematic knowledge follow a similar time course, from around 750ms after target-word onset. The analyses suggest that

toddlers demonstrated a stronger preference for thematic distractors than taxonomic distractors. Furthermore, we found that expressive vocabulary size predicted thematic fixations while neither receptive nor expressive vocabulary size were positively associated with taxonomic fixations. Perhaps, expressive vocabulary size is an indicator of the number of opportunities to learn to associate two thematically-related words in the lexical-semantic system. However, such a comparison must be taken with caution as the thematic distractor was designed to be also associated with the target, while the taxonomic distractor was not. Previous priming studies have shown that added association boosts the size of semantic priming effects (Moss et al., 1995). Our findings of a weaker taxonomic effect and a stronger thematic/associated effect could be explained by Plaut and Booth's (2000) model. According to their model, associative priming occurs because the lexical-semantic system has learned to associate the occurrence of one word/concept with another, and as a result, hearing one word leads to the automatic activation of the other. In contrast, taxonomic priming occurs because the prime word activates features that overlap with taxonomically-related concepts. Taxonomic priming effects tend to be less strong because non-shared features between the prime and target leads to inhibition. For example, bees have wings while cats do not.

General Discussion

Using the four-picture adult visual world paradigm (VWP), this study examined the mental processes involved in lexical access in toddlers between 24- to 30-months. Across two experiments we have demonstrated that language can direct toddlers' attention to related phonological and semantic pictures amongst an array of four pictures in less than 1000ms. The speed with which toddlers are able to launch saccades and land eye fixations on pictures that are named or subtly related to a complex auditory stimulus attests to a high degree of sophistication

in the integration of meaningful auditory and visual signals. More importantly, such language-mediated attention is influenced by both the phonological and semantic properties associated with the visual stimuli. Pictures whose labels share the same onset consonant as the spoken target word attract attention faster than pictures that share semantic features with the referent of the spoken word. The time course and order of the phonological and semantic preferences likely reflect the automatic activation of phonological and semantic representations during spoken-word recognition in toddlers. These findings resemble those reported for adults (Huettig & McQueen, 2007) and extend them to a much lower age (24–30 months), suggesting important representational properties of the adult lexical-semantic system, such as information cascade, are active during early lexical development.

Taxonomic and Thematic

We separately examined the extraction of two types of semantic relations during spoken-word recognition. Upon hearing a label, toddlers showed increased preference for an object that belongs to the same taxonomic category as the label (e.g., bee and cat) (Experiment 1) or one that belongs to the same thematic category as the label (e.g., brush and hair) (Experiment 2). These results suggest that the toddler's lexical-semantic system encodes both taxonomic and thematic relations between words, and that spoken-word recognition not only involves the matching of the label and its meaning, but also the activation of taxonomically- and thematically-related ones. We compared the time course and magnitude of toddlers' preference for the taxonomic and thematic distractors. The time course did not differ between the two types of distractors. However, toddlers' preference for the thematic distractor was substantially stronger than that for the taxonomic distractor. While it is tempting to conclude that the strength of thematic/associative relations are stronger than that of purely taxonomic relations (see

Discussion of Experiment 2 on the ‘association boost’), it is possibly more constructive to enquire as to the extent these two types of semantic relations are similar or different from each other.

Taxonomic relations are typically defined in terms of shared functional and visual features. In our study, we used taxonomic distractors that are unassociated to the target label to ensure that the source of any taxonomic preference is purely taxonomic in nature, rather than a combination of both associative and taxonomic. Unassociated taxonomic pairs tend not to share visual features (e.g., bee and cat, compared to associated taxonomic pairs such as dog and cat). Therefore, while the label ‘bee’ activates cat because both share the features of living things, the non-shared features of bee (e.g., wings and a sting) inhibits the activation of cat, resulting in reduction of the net activation level (e.g., Plaut & Booth, 2002). Thematic relations are not typically defined by shared features, but by the correlation of feature co-occurrences. For example, brush and hair have little to no shared features. Instead, we suppose that it is the association or the encoded co-occurrence of the features of brush and hair in the toddler’s lexical semantic system that leads to their mutual activation during spoken-word recognition. It is also possible that toddlers’ thematic preference was driven directly by the lexical association between the target label and the thematic distractor (Plaut & Booth, 2002). Associative learning may prioritise the formation of thematic links over the more analytic, feature-based taxonomic links in the early developing lexical-semantic system, perhaps leading to the stronger thematic preferences reported in the current study.

Vocabulary Size and Language-mediated Visual Attention

In Experiment 1, we found that toddlers with a larger receptive vocabulary showed a greater preference for the phonological distractor than those with a smaller receptive vocabulary.

As a result, toddlers with a larger receptive vocabulary began to fixate the taxonomic distractor later in the time course, showing an overall smaller preference for the taxonomic distractor in the window of analysis. In Experiment 2, we tested another group of toddlers and found that those with a larger expressive vocabulary showed a greater preference for both the phonological and thematic distractors than those with a smaller expressive vocabulary. We do not fully understand why receptive vocabulary but not expressive vocabulary predicts performance in Experiment 1, and vice versa in Experiment 2. Nevertheless, our findings are consistent with the view that infants with larger vocabularies are likely to have more refined and consolidated phonological representations which in turn impacts the activation of phonologically related words, and hence the identification and fixation of phonologically related competitors.

Nor is it entirely clear why toddlers with a larger vocabulary showed greater preference for the thematic distractor, but not for the taxonomic distractor. As discussed in the previous section, the encoding and retrieval of taxonomic and thematic relations in the lexical semantic system are likely different. The thematic distractor was chosen to be associated with the target word based on adult word association norms. A larger expressive vocabulary may reflect greater opportunities to utter thematically-related words, which in turn strengthen their association in the lexical-semantic system (see Discussion of Experiment 2). In contrast, the taxonomic distractor was designed to be unassociated with the target, and that most of them have shared as well as non-shared features with the target. Without further evidence, it is impossible to pinpoint whether the lack of positive relationship between vocabulary size and taxonomic preference is simply due to the lack of association, or that the non-shared features cancel out any potential effect of vocabulary size, or both.

Nonetheless, our findings of a positive relationship between vocabulary size and language processing is not a standalone case in the literature. Our findings corroborate previous IPL studies showing that toddlers with a larger vocabulary can recognise spoken-words faster (e.g., Fernald et al., 2006; Fernald et al., 2001; Zangl et al., 2005), indicating that the more words the toddlers know or say, the better they are at selectively attending to relevant information. The reverse is likely to be also true: the better toddlers are at selectively attending to relevant information, the faster and the more words they learn. Some studies have demonstrated that toddlers' vocabulary size predicts academic performance or language learning outcomes later in life (e.g., Fernald et al., 2006; Marchman & Fernald, 2008; Morgan, Farkas, Hillemeier, Hammer & Maczuga, 2015), and that children from low social economic status households tend to have a lower vocabulary during early childhood (e.g., Fernald, Marchman & Weisleder, 2012; Lee & Burkam, 2002; Morgan et al., 2015) or language delays (Nelson, Welsh, Vance Trup & Greenberg, 2011). Future research should endeavour to disentangle the relationship between vocabulary size, intelligence, language-mediated visual attention and subsequently learning outcomes, possibly involving the use of longitudinal designs.

Phonological vs. Semantic

In both Experiments 1 and 2, the named target referents were absent from the visual array for the critical test trials. A target-absent scenario is important because it permits potential relationships between the spoken word and the different types of distractors to become apparent, unmasked by the attention attracting impact of the target itself. In both target-absent experiments, toddlers began to demonstrate an early preference for the phonological distractor over the unrelated distractor from around 300ms to 700ms after target-word onset. However, fixation of the phonological distractor was short-lived. Toddlers' preference for the phonological

distractor resumed to baseline level within approximately 1000ms after target word onset. As attention towards the phonological distractor reduced, toddlers were drawn to the semantic distractor (either taxonomic or thematic) from around 750ms after target word onset. Based on previous adult VWP studies that used picture stimuli, the phonological and semantic effects observed in this study were likely driven by the match of activated information at the phonological and semantic levels respectively (Huetting & McQueen, 2007). These results suggest that toddlers, like adults, extract phonological and semantic information from speech in an incremental fashion, leading to the cascaded activation of unnamed but phonologically- and semantically-related items in the lexical-semantic system.

The overall effect of phonological activation was smaller than that for taxonomic or thematic activation (see Figure 8). This may be due to the fact that the taxonomic/thematic distractors were inherently more related to the target than the phonological distractor. As the cohort model predicts, candidate words that no longer match the target word become irrelevant as the target word unfolds (Marslen-Wilson, 1987). In contrast, the taxonomic and thematic distractors should become increasingly salient after the target word has fully unfolded, particularly in the absence of the named target.

We have assumed that phonological and semantic knowledge are represented at different levels in the lexical-semantic system, and that access of knowledge at the phonological level precedes the semantic level (e.g., Marslen-Wilson, 1987; McClelland & Elman, 1986). However, the different time course of phonological and semantic activation can also be explained by a distributed model in which phonological and semantic features are represented at the same level and can be activated simultaneously (Gaskell & Marslen-Wilson, 1997, 2002). For example, hearing the target word would lead to the activation of phonological and semantic features in

parallel. Since speech is processed incrementally, the activated phonological features tend to be more informative early in the time course. In contrast, semantic features would receive relatively weaker activation and ambiguous activation early in the time course, but become more informative later in the time course as the target word fully unfolds. Our results do not adjudicate between a multilevel lexical-semantic system and a distributed feature approach, since both predict the cascaded activation from phonological to semantic features observed in our experiment.

Identifying the Locus of the Phonological and Semantic Effects

We now reconsider how the phonological, semantic and visual hypotheses may explain our findings of an earlier phonological preference and a later semantic preference (see the Introduction for a brief description of these hypotheses). To facilitate explication, Figure 9 visualises the possible ‘steps’¹⁴ required to activate and match the different types of representations in the auditory- and visual-input routes. As the time course of the phonological and semantic preference is driven by the onset of the target label, we will count the steps in the auditory-input route only. The phonological and semantic hypotheses assume that participants’ preference for the phonological and semantic distractors are driven by a match at the phonological and semantic representation level, respectively. It is important to note that a match at the phonological and semantic levels is only possible if it is assumed that viewing of pictures leads to automatic cascaded activation of the picture’s semantic and phonological information. As illustrated in Figure 9a, the earlier phonological preference and a later semantic preference can be explained by a phonological match (one step of activation) followed by a semantic match (three steps of activation).

¹⁴ This is by no means an exhaustive illustration of all possible ‘steps’ of lexical-semantic activation.

In contrast, the visual hypothesis (c.f. Dahan & Tanenhaus, 2005) maintains that language-mediated visual attention is driven by a visual match, and denies the automatic activation of phonological representations when viewing a picture. Based on these assumptions, viewers are effectively ‘blind’ to the phonological and semantic relations between the label and the pictures until the relevant visual-form representations have been activated. Our findings can be explained by the visual hypothesis under the assumption that cascaded activation of the visual form of the phonological distractor is achieved faster than that of the semantic distractor. As illustrated in Figure 9b, it would take three steps of activation for the phonological distractor (house) to be matched on the visual level, and four steps of activation for the thematic distractor (brush). Compared to a phonological and semantic match, a visual match requires more ‘steps’ or time to drive visual attention to the thematic distractor.

Figure 9. Visualisation of the possible ‘steps’ of activation and level of matching based on the (a) phonological and semantic hypotheses and (b) the visual hypothesis.

Although the results of our experiments are consistent with either of these accounts, we note that the phonological and visual hypotheses need not be mutually exclusive. Adults and toddlers in these experiments have both auditory and visual information available to them during the course of a trial, and often the visual stimuli are available prior to the onset of any auditory stimulus. Given that experimental participants expect to hear a name during a trial, it is not implausible to suppose that some kind of verbal anticipation plays a role in deciding which picture to fixate. We might therefore expect a complex interplay between the activity of representations resulting from cascading information from the visual input and that from the auditory input. Furthermore, the speed, magnitude and duration with which different types of representation become active will no doubt depend on task factors, such as the relative timing of visual and auditory input, and the nature and quality of the stimuli themselves. For example, line drawings, photographs and orthographic presentations may differentially impact the extraction of phonological codes from the visual stimulus and thereby the locus of the match which supports participants' picture fixations. Clearly, further research is required to achieve a fuller understanding of how language drives attention in a visual world.

In an article titled 'Why we're so smart?', Gentner (2003) proposed that, together with a sophisticated learning system, human language acts as a tool to structure knowledge by encoding the similarities and dissimilarities between objects and concepts. This knowledge in turn provides the language user relational insights in novel situations, and assists children to learn abstract knowledge, make inferences and engage in analogical reasoning. Our findings underpin Gentner's (2003) proposal: During spoken-word recognition, phonological, taxonomic and thematic codes are automatically activated to bias our selective attention to relevant information in the visual world, which is likely to enhance the efficiency of mental activities, such as learning

and making analogical inferences.

References

- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, 38(4), 419–439.
- Annett, M. (1959). The classification of instances of four common class concepts by children and adults. *British Journal of Educational Psychology*, 29(3), 223-236.
- Arias-Trejo, N., & Plunkett, K. (2009). Lexical–semantic priming effects during infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1536), 3633–3647.
- Arias-Trejo, N., & Plunkett, K. (2013). What’s in a link: Associative and taxonomic priming effects in the infant lexicon. *Cognition*, 128(2), 214–227.
- Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, 17(2), 1265-1282.
- Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1; 2. *Journal of Child Language*, 32(01), 159-173.
- Barr, D. J. (2008). Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of memory and language*, 59(4), 457–474.
- Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar*. Cambridge University Press.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10, 433–436.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing.

Psychological review, 82(6), 407.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6(1), 84–107.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, 132(2), 163.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic bulletin & review*, 12(3), 453-459.

Denney, N. W. (1974). Evidence for developmental changes in categorization criteria for children and adults. *Human Development*, 17(1), 41-53.

Estes, Z., Golonka, S., & Jones, L. L. (2011). 8 Thematic Thinking: The Apprehension and Consequences of Thematic Relations. *Psychology of Learning and Motivation-Advances in Research and Theory*, 54, 249.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental science*, 16(2), 234-248.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental*

psychology, 42(1), 98.

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child development*, 72(4), 1003–1015.

Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of cognitive neuroscience*, 16(8), 1465-1477.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and cognitive Processes*, 12(5-6), 613–656.

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive psychology*, 45(2), 220–266.

Gentner, D. (2003). Why we're so smart. In D. Gentner and S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought*. Cambridge (pp.195-235), MA: MIT Press.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of child language*, 14(01), 23–45.

Greenfield, D. B., & Scott, M. S. (1986). Young children's preference for complementary pairs: Evidence against a shift to a taxonomic preference. *Developmental Psychology*, 22(1), 19.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a

- british communicative development inventory. *Journal of child language*, 27(03), 689–705.
- Huetting, F., & Altmann, G. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, 96(1), B23–B32.
- Huetting, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, 57(4), 460–482.
- Huetting, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171.
- Inhelder, B., & Piaget, J. (1964). The early growth of logic in the child: Classification and seriation (EA Lunzer & D. Papert, Trans.), London: Routledge and Kegan Paul.
- Johnson, E. K., McQueen, J. M., & Huetting, F. (2011). Toddlers' language-mediated visual search: They need not have the words for it. *The Quarterly Journal of Experimental Psychology*, 64(9), 1672–1682.
- Kiss, G., Armstrong, C. A., & Milroy, R. (1972). *An associative thesaurus of English*. Medical Research Council, Speech and Communication Unit, University of Edinburgh, Scotland.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, 36(14), 1.
- Law II, F., & Edwards, J. R. (2014). Effects of vocabulary size on online lexical processing by

- preschoolers. *Language Learning and Development*, 1–25.
- Lee, V. E., & Burkham, D. T. (2002). Inequality at the starting gate: Social background differences in achievement as children begin Kindergarten. *Washington, DC: Economic Policy Institute*.
- Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, 130(1), 3.
- Mani, N., Durrant, S., & Floccia, C. (2012). Activation of phonological and semantic codes in toddlers. *Journal of Memory and Language*, 66(4), 612–622.
- Mani, N., & Plunkett, K. (2010). In the infant's mind's ear evidence for implicit naming in 18-month-olds. *Psychological science*, 21(7), 908–913.
- Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental science*, 11(3), F9-F16.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1), 71–102.
- Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human perception and performance*, 15(3), 576.
- Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from cdi analysis. *Developmental Science*, 14(4), 769–785.

- McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive psychology*, 18(1), 1–86.
- McQueen, J. M., & Huettig, F. (2014). Interference of spoken word recognition through phonological priming from visual objects and printed words. *Attention, Perception, & Psychophysics*, 76(1), 190-200.
- Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual review of psychology*, 35(1), 113–138.
- Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. Ehri (Eds.), (p. 89-120). Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.
- Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review*, 14(4), 710-716.
- Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & cognition*, 35(3), 494–503.
- Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 555.
- Mirman, D. (2014). *Growth curve analysis and visualization using r*. Taylor & Francis.
- Mirman, D., & Graziano, K. M. (2012). Individual differences in the strength of taxonomic

- versus thematic relations. *Journal of experimental psychology: General*, 141(4), 601.
- Moss, H., Hare, M., Day, P., & Tyler, L. (1994). A distributed memory model of the associative boost in semantic priming. *Connection Science*, 6(4), 413–427.
- Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-Month-Old Children With Larger Oral Vocabularies Display Greater Academic and Behavioral Functioning at Kindergarten Entry. *Child development*, 86(5), 1351-1370.
- Moss, H. E., Hare, M. L., Day, P., & Tyler, L. K. (1994). A distributed memory model of the associative boost in semantic priming. *Connection Science*, 6(4), 413-427.
- Moss, H., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 863.
- Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, 53(3), 359-377.
- Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the society for research in child development*, 1-135.
- Nelson, D., McEvoy, C., & Schreiber, T. (1998). The university of south florida word association, rhyme, and word fragment norms. *Free Association*. Retrieved from <http://www.usf.edu/FreeAssociation/>.
- Nelson, K. E., Welsh, J. A., Trup, E. M. V., & Greenberg, M. T. (2011). Language delays of impoverished preschool children in relation to early academic and emotion recognition

- skills. *First Language*, 0142723710391887.
- Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10(4), 437–442.
- Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: empirical and computational support for a single-mechanism account of lexical processing. *Psychological review*, 107(4), 786.
- Posnansky, C. J. (1978). Category norms for verbal items in 25 categories for children in Grades 2–6. *Behavior Research Methods & Instrumentation*, 10(6), 819-832.
- Rämä, P., Sirri, L., & Serres, J. (2013). Development of lexical–semantic language system: N400 priming effect for spoken words in 18-and 24-month old children. *Brain and language*, 125(1), 1-10.
- Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1108.
- Scott, M. S., Serchuk, R., & Mundy, P. (1982). Taxonomic and complementary picture pairs: Ability in two to five-year-olds. *International Journal of Behavioral Development*, 5(2), 243–256.
- Smiley, S. S., & Brown, A. L. (1979). Conceptual preference for thematic or taxonomic relations: A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology*, 28(2), 249-257.

- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Harvard University Press
Cambridge, MA.
- Strijkers, K., Holcomb, P. J., & Costa, A. (2011). Conscious intention to speak proactively facilitates lexical access during overt object naming. *Journal of Memory and Language*, 65(4), 345-362.
- Styles, S. J., & Plunkett, K. (2009). How do infants build a semantic system? *Language and Cognition*, 1(1), 1-24.
- Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76(2), 147-166.
- Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, 71(2), 73-108.
- Tyler, L. K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in cognitive sciences*, 5(6), 244-252.
- Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of experimental child psychology*, 113(4), 569-586.
- Waxman, S. R., & Namy, L. L. (1997). Challenging the notion of a thematic preference in young children. *Developmental Psychology*, 33(3), 555.
- Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, 3(1), 1-30.
- Yee, E., Overton, E., & Thompson-Schill, S. L. (2009). Looking for meaning: Eye movements

are sensitive to overlapping semantic features, not association. *Psychonomic bulletin & review*, 16(5), 869–874.

Zangl, R., Klarman, L., Thal, D., Fernald, A., & Bates, E. (2005). Dynamics of word comprehension in infancy: Developments in timing, accuracy, and resistance to acoustic degradation. *Journal of Cognition and Development*, 6(2), 179–208.

Appendix A

Table A1

Experiment 1 stimuli yoke list.

Yoke Set	Target Word	Distractor Pictures			
		<i>Phonological</i>	<i>Taxonomic</i>	<i>Unrelated 1</i>	<i>Unrelated 2</i>
1	Trousers	Train	Hat	Fridge	Bear
	Boat	Bear	Train	Hat	Fridge
	Hoover	Hat	Fridge	Bear	Train
	Frog	Fridge	Bear	Train	Hat
2	Squirrel	Soap	Duck	Blocks	Tomato
	Doll	Duck	Blocks	Tomato	Soap
	Towel	Tomato	Soap	Duck	Blocks
	Banana	Blocks	Tomato	Soap	Duck
3	Boots	Butterfly	Coat	Peas	Tractor
	Tiger	Tractor	Butterfly	Coat	Peas
	Cake	Coat	Peas	Tractor	Butterfly
	Pushchair	Peas	Tractor	Butterfly	Coat
4	Digger	Dress	Bus	Cat	Sandwich
	Cucumber	Cat	Sandwich	Dress	Bus
	Bee	Bus	Cat	Sandwich	Dress
	Sock	Sandwich	Dress	Bus	Cat
5	Bath	Bicycle	Sofa	Carrot	Turtle
	Cow	Carrot	Turtle	Bicycle	Sofa
	Strawberry	Sofa	Carrot	Turtle	Bicycle
	Truck	Turtle	Bicycle	Sofa	Carrot
6	Toast	TV	Pear	Horse	Car
	Helicopter	Horse	Car	TV	Pear
	Cot	Car	TV	Pear	Horse
	Penguin	Pear	Horse	Car	TV

Table A2

Experiment 1–Test List 1.

List 1			Picture Position			
			Upper Left	Upper Right	Lower Left	Lower Right
Block	Type	Target	Pictures (Picture Role)			
0	Filler	Juice	Juice (t)	Nappy (u1)	Brush (u2)	Keys (u3)
	Filler	Fire-engine	Grape (u)	Fire-engine (t)	Fish (p)	Aeroplane (c)
	Test	Trousers	Train (p)	Hat (c)	Bear (u2)	Fridge (u1)
	Test	Squirrel	Duck (c)	Soap (p)	Blocks (u1)	Tomato (u2)
	Filler	Chocolate	Icecream (c)	Umbrella (u)	Chocolate (t)	Chicken (p)
	Test	Bath	Bicycle (p)	Carrot (u1)	Sofa (c)	Turtle (u2)
	Test	Toast	Pear (c)	Car (u2)	TV (p)	Horse (u1)
	Filler	Bib	Book (p)	Shoes (c)	Clock (u)	Bib (t)
1	Test	Digger	Sandwich (u2)	Dress (p)	Cat (u1)	Bus (c)
	Test	Boots	Peas (u1)	Butterfly (p)	Coat (c)	Tractor (u2)
	Filler	Monkey	Monkey (t)	Milk (p)	Dog (c)	Telephone (u)
	Test	Boat	Hat (u1)	Bear (p)	Fridge (u2)	Train(c)
	Test	Cow	Sofa (u2)	Turtle (c)	Bicycle (u1)	Carrot (p)
	Filler	Spoon	Ambulance (u)	Spoon (t)	Swing (p)	Glass (c)
	Test	Tiger	Butterfly (c)	Coat (u1)	Tractor (p)	Peas (u2)
	Test	Cucumber	Cat (p)	Bus (u2)	Sandwich (c)	Dress (u1)
2	Filler	Teddy	Ball (c)	Flower (u)	Teddy (t)	Table (p)
	Test	Helicopter	TV (u1)	Horse (p)	Pear (u2)	Car (c)
	Test	Doll	Soap (u2)	Blocks (c)	Tomato (u1)	Duck (p)
	Filler	Pig	Pen (p)	Elephant (c)	Button (u)	Pig (t)
	Test	Strawberry	Turtle (u1)	Bicycle (u2)	Carrot (c)	Sofa (p)
	Test	Bee	Dress (u2)	Cat (c)	Bus (p)	Sandwich (u1)
	Filler	Hand	Hand (t)	Raisins (u1)	Tree (u2)	Window (u3)
	Test	Cake	Tractor (u1)	Peas (c)	Butterfly (u2)	Coat (p)
3	Test	Hoover	Fridge (c)	Train (u2)	Hat (p)	Bear (u1)
	Filler	Cup	Watch (u)	Cup (t)	Caterpillar (p)	Fork (c)
	Test	Towel	Tomato (p)	Duck (u1)	Soap (c)	Blocks (u2)
	Test	Cot	Car (p)	TV (c)	Horse (u2)	Pear (u1)
	Filler	Spade	Watering Can (c)	Door (u)	Spade (t)	Snail (p)
	Test	Pushchair	Coat (u2)	Tractor (c)	Peas (p)	Butterfly (u1)
	Test	Sock	Bus (u1)	Sandwich (p)	Dress (c)	Cat (u2)
	Filler	Bunny	Bag (p)	Lion (c)	Plate (u)	Bunny (t)
4	Test	Truck	Carrot (u2)	Sofa (u1)	Turtle (p)	Bicycle (c)
	Test	Frog	Bear (c)	Fridge (p)	Train (u1)	Hat (u2)
	Filler	Slide	Slide (t)	Cooker (u1)	Orange (u2)	Toothbrush (u3)
	Test	Banana	Blocks (p)	Tomato (c)	Duck (u2)	Soap (u1)
	Test	Penguin	Horse (c)	Pear (p)	Car (u1)	TV (u2)

Note: p = phonological distractor, c = taxonomic distractor, u = unrelated distractors, t = target

Table A3

Experiment 1–Test List 2.

List 2			Picture Position			
			Upper Left	Upper Right	Lower Left	Lower Right
Block	Type	Target	Pictures (Picture Role)			
0	Filler	Juice	Nappy (u1)	Brush (u2)	Keys (u3)	Juice (t)
	Filler	Monkey	Milk (p)	Dog (c)	Telephone (u)	Monkey (t)
	Test	Boat	Bear (p)	Fridge (u2)	Train(c)	Hat (u1)
	Test	Cow	Turtle (c)	Bicycle (u1)	Carrot (p)	Sofa (u2)
	Filler	Spoon	Spoon (t)	Swing (p)	Glass (c)	Ambulance (u)
	Test	Tiger	Coat (u1)	Tractor (p)	Peas (u2)	Butterfly (c)
	Test	Cucumber	Bus (u2)	Sandwich (c)	Dress (u1)	Cat (p)
	Filler	Teddy	Flower (u)	Teddy (t)	Table (p)	Ball (c)
2	Test	Helicopter	Horse (p)	Pear (u2)	Car (c)	TV (u1)
	Test	Doll	Blocks (c)	Tomato (u1)	Duck (p)	Soap (u2)
	Filler	Pig	Elephant (c)	Button (u)	Pig (t)	Pen (p)
	Test	Strawberry	Bicycle (u2)	Carrot (c)	Sofa (p)	Turtle (u1)
	Test	Bee	Cat (c)	Bus (p)	Sandwich (u1)	Dress (u2)
	Filler	Hand	Raisins (u1)	Tree (u2)	Window (u3)	Hand (t)
	Test	Cake	Peas (c)	Butterfly (u2)	Coat (p)	Tractor (u1)
	Test	Hoover	Train (u2)	Hat (p)	Bear (u1)	Fridge (c)
3	Filler	Cup	Cup (t)	Caterpillar (p)	Fork (c)	Watch (u)
	Test	Towel	Duck (u1)	Soap (c)	Blocks (u2)	Tomato (p)
	Test	Cot	TV (c)	Horse (u2)	Pear (u1)	Car (p)
	Filler	Spade	Door (u)	Spade (t)	Snail (p)	Watering Can (c)
	Test	Pushchair	Tractor (c)	Peas (p)	Butterfly (u1)	Coat (u2)
	Test	Sock	Sandwich (p)	Dress (c)	Cat (u2)	Bus (u1)
	Filler	Bunny	Lion (c)	Plate (u)	Bunny (t)	Bag (p)
	Test	Truck	Sofa (u1)	Turtle (p)	Bicycle (c)	Carrot (u2)
4	Test	Frog	Fridge (p)	Train (u1)	Hat (u2)	Bear (c)
	Filler	Slide	Cooker (u1)	Orange (u2)	Toothbrush (u3)	Slide (t)
	Test	Banana	Tomato (c)	Duck (u2)	Soap (u1)	Blocks (p)
	Test	Penguin	Pear (p)	Car (u1)	TV (u2)	Horse (c)
	Filler	Fire-engine	Fire-engine (t)	Fish (p)	Aeroplane (c)	Grape (u)
	Test	Trousers	Hat (c)	Bear (u2)	Fridge (u1)	Train (p)
	Test	Squirrel	Soap (p)	Blocks (u1)	Tomato (u2)	Duck (c)
	Filler	Chocolate	Umbrella (u)	Chocolate (t)	Chicken (p)	Icecream (c)
1	Test	Bath	Carrot (u1)	Sofa (c)	Turtle (u2)	Bicycle (p)
	Test	Toast	Car (u2)	TV (p)	Horse (u1)	Pear (c)
	Filler	Bib	Shoes (c)	Clock (u)	Bib (t)	Book (p)
	Test	Digger	Dress (p)	Cat (u1)	Bus (c)	Sandwich (u2)
	Test	Boots	Butterfly (p)	Coat (c)	Tractor (u2)	Peas (u1)

Note: p = phonological distractor, c = taxonomic distractor, u = unrelated distractors, t = target

Table A4

Experiment 1–Test List 3.

List 3			Picture Position			
			Upper Left	Upper Right	Lower Left	Lower Right
Block	Type	Target	Pictures (Picture Role)			
0	Filler	Juice	Brush (u2)	Keys (u3)	Juice (t)	Nappy (u1)
	Filler	Pig	Button (u)	Pig (t)	Pen (p)	Elephant (c)
	Test	Strawberry	Carrot (c)	Sofa (p)	Turtle (u1)	Bicycle (u2)
	Test	Bee	Bus (p)	Sandwich (u1)	Dress (u2)	Cat (c)
	Filler	Hand	Tree (u2)	Window (u3)	Hand (t)	Raisins (u1)
	Test	Cake	Butterfly (u2)	Coat (p)	Tractor (u1)	Peas (c)
	Test	Hoover	Hat (p)	Bear (u1)	Fridge (c)	Train (u2)
	Filler	Cup	Caterpillar (p)	Fork (c)	Watch (u)	Cup (t)
	Test	Towel	Soap (c)	Blocks (u2)	Tomato (p)	Duck (u1)
3	Test	Cot	Horse (u2)	Pear (u1)	Car (p)	TV (c)
	Filler	Spade	Spade (t)	Snail (p)	Watering Can (c)	Door (u)
	Test	Pushchair	Peas (p)	Butterfly (u1)	Coat (u2)	Tractor (c)
	Test	Sock	Dress (c)	Cat (u2)	Bus (u1)	Sandwich (p)
	Filler	Bunny	Plate (u)	Bunny (t)	Bag (p)	Lion (c)
	Test	Truck	Turtle (p)	Bicycle (c)	Carrot (u2)	Sofa (u1)
	Test	Frog	Train (u1)	Hat (u2)	Bear (c)	Fridge (p)
	Filler	Slide	Orange (u2)	Toothbrush (u3)	Slide (t)	Cooker (u1)
	Test	Banana	Duck (u2)	Soap (u1)	Blocks (p)	Tomato (c)
4	Test	Penguin	Car (u1)	TV (u2)	Horse (c)	Pear (p)
	Filler	Fire-engine	Fish (p)	Aeroplane (c)	Grape (u)	Fire-engine (t)
	Test	Trousers	Bear (u2)	Fridge (u1)	Train (p)	Hat (c)
	Test	Squirrel	Blocks (u1)	Tomato (u2)	Duck (c)	Soap (p)
	Filler	Chocolate	Chocolate (t)	Chicken (p)	Icecream (c)	Umbrella (u)
	Test	Bath	Sofa (c)	Turtle (u2)	Bicycle (p)	Carrot (u1)
	Test	Toast	TV (p)	Horse (u1)	Pear (c)	Car (u2)
	Filler	Bib	Clock (u)	Bib (t)	Book (p)	Shoes (c)
	Test	Digger	Cat (u1)	Bus (c)	Sandwich (u2)	Dress (p)
1	Test	Boots	Coat (c)	Tractor (u2)	Peas (u1)	Butterfly (p)
	Filler	Monkey	Dog (c)	Telephone (u)	Monkey (t)	Milk (p)
	Test	Boat	Fridge (u2)	Train(c)	Hat (u1)	Bear (p)
	Test	Cow	Bicycle (u1)	Carrot (p)	Sofa (u2)	Turtle (c)
	Filler	Spoon	Swing (p)	Glass (c)	Ambulance (u)	Spoon (t)
	Test	Tiger	Tractor (p)	Peas (u2)	Butterfly (c)	Coat (u1)
	Test	Cucumber	Sandwich (c)	Dress (u1)	Cat (p)	Bus (u2)
	Filler	Teddy	Teddy (t)	Table (p)	Ball (c)	Flower (u)
	Test	Helicopter	Pear (u2)	Car (c)	TV (u1)	Horse (p)
2	Test	Doll	Tomato (u1)	Duck (p)	Soap (u2)	Blocks (c)

Note: p = phonological distractor, c = taxonomic distractor, u = unrelated distractors, t = target

Table A5

Experiment 1–Test List 4.

List 4			Picture Position			
			<i>Upper Left</i>	<i>Upper Right</i>	<i>Lower Left</i>	<i>Lower Right</i>
Block	Type	Target	Pictures (Picture Role)			
0	Filler	Juice	Keys (u3)	Juice (t)	Nappy (u1)	Brush (u2)
	Filler	Spade	Snail (p)	Watering Can (c)	Door (u)	Spade (t)
	Test	Pushchair	Butterfly (u1)	Coat (u2)	Tractor (c)	Peas (p)
	Test	Sock	Cat (u2)	Bus (u1)	Sandwich (p)	Dress (c)
	Filler	Bunny	Bunny (t)	Bag (p)	Lion (c)	Plate (u)
	Test	Truck	Bicycle (c)	Carrot (u2)	Sofa (u1)	Turtle (p)
	Test	Frog	Hat (u2)	Bear (c)	Fridge (p)	Train (u1)
	Filler	Slide	Toothbrush (u3)	Slide (t)	Cooker (u1)	Orange (u2)
	Test	Banana	Soap (u1)	Blocks (p)	Tomato (c)	Duck (u2)
1	Test	Penguin	TV (u2)	Horse (c)	Pear (p)	Car (u1)
	Filler	Fire-engine	Aeroplane (c)	Grape (u)	Fire-engine (t)	Fish (p)
	Test	Trousers	Fridge (u1)	Train (p)	Hat (c)	Bear (u2)
	Test	Squirrel	Tomato (u2)	Duck (c)	Soap (p)	Blocks (u1)
	Filler	Chocolate	Chicken (p)	Icecream (c)	Umbrella (u)	Chocolate (t)
	Test	Bath	Turtle (u2)	Bicycle (p)	Carrot (u1)	Sofa (c)
	Test	Toast	Horse (u1)	Pear (c)	Car (u2)	TV (p)
	Filler	Bib	Bib (t)	Book (p)	Shoes (c)	Clock (u)
	Test	Digger	Bus (c)	Sandwich (u2)	Dress (p)	Cat (u1)
2	Test	Boots	Tractor (u2)	Peas (u1)	Butterfly (p)	Coat (c)
	Filler	Monkey	Telephone (u)	Monkey (t)	Milk (p)	Dog (c)
	Test	Boat	Train(c)	Hat (u1)	Bear (p)	Fridge (u2)
	Test	Cow	Carrot (p)	Sofa (u2)	Turtle (c)	Bicycle (u1)
	Filler	Spoon	Glass (c)	Ambulance (u)	Spoon (t)	Swing (p)
	Test	Tiger	Peas (u2)	Butterfly (c)	Coat (u1)	Tractor (p)
	Test	Cucumber	Dress (u1)	Cat (p)	Bus (u2)	Sandwich (c)
	Filler	Teddy	Table (p)	Ball (c)	Flower (u)	Teddy (t)
	Test	Helicopter	Car (c)	TV (u1)	Horse (p)	Pear (u2)
3	Test	Doll	Duck (p)	Soap (u2)	Blocks (c)	Tomato (u1)
	Filler	Pig	Pig (t)	Pen (p)	Elephant (c)	Button (u)
	Test	Strawberry	Sofa (p)	Turtle (u1)	Bicycle (u2)	Carrot (c)
	Test	Bee	Sandwich (u1)	Dress (u2)	Cat (c)	Bus (p)
	Filler	Hand	Window (u3)	Hand (t)	Raisins (u1)	Tree (u2)
	Test	Cake	Coat (p)	Tractor (u1)	Peas (c)	Butterfly (u2)
	Test	Hoover	Bear (u1)	Fridge (c)	Train (u2)	Hat (p)
	Filler	Cup	Fork (c)	Watch (u)	Cup (t)	Caterpillar (p)
	Test	Towel	Blocks (u2)	Tomato (p)	Duck (u1)	Soap (c)
	Test	Cot	Pear (u1)	Car (p)	TV (c)	Horse (u2)

Note: p = phonological distractor, c = taxonomic distractor, u = unrelated distractors, t = target

Appendix B

Table B1

Experiment 2 stimuli yoke list.

Yoke Set	Target Word	Distractor Pictures			
		<i>Phonological</i>	<i>Thematic</i>	<i>Unrelated 1</i>	<i>Unrelated 2</i>
1	Sky	Spade	Rocket	Windows	Carrot
	Rabbit	Rocket	Carrot	Spade	Windows
	Curtains	Carrot	Windows	Rocket	Spade
	Watering Can	Windows	Spade	Carrot	Rocket
2	Hair	House	Brush	Flower	Camera
	Key	Camera	House	Brush	Flower
	Butterfly	Brush	Flower	Camera	House
	Photo	Flower	Camera	House	Brush
3	Banana	Bin	Monkey	Cup	Road
	Milk	Monkey	Cup	Road	Bin
	Rubbish	Road	Bin	Monkey	Cup
	Car	Cup	Road	Bin	Monkey
4	Party	Papers	Balloon	Tree	Spoon
	Book	Balloon	Papers	Spoon	Tree
	Squirrel	Spoon	Tree	Papers	Balloon
	Table	Tree	Spoon	Balloon	Papers
5	Shoe	Sheep	Foot	Bowl	Soap
	Farm	Foot	Sheep	Soap	Bowl
	Cereal	Soap	Bowl	Foot	Sheep
	Bath	Bowl	Soap	Sheep	Foot
6	Baby	Button	Pushchair	Cushions	Swing
	Park	Pushchair	Swing	Button	Cushions
	Sofa	Swing	Cushions	Pushchair	Button
	Coat	Cushions	Button	Swing	Pushchair

Table B2

Experiment 2–Test List 1.

List 1			Picture Position			
			Upper Left	Upper Right	Lower Left	Lower Right
Block	Type	Target	Pictures (Picture Role)			
0	Filler	grape	grape (t)	highchair (u1)	train (u2)	fish (u3)
	Filler	sock	pear (u3)	sock (t)	tV (u1)	bicycle (u2)
	Test	sky	spade (p)	rocket (a)	windows (u1)	carrot (u2)
	Test	butterfly	flower (a)	camera (u1)	house (u2)	brush (p)
	Filler	chicken	fork (u2)	bus (u3)	chicken (t)	nappy (u1)
	Test	rubbish	cup (u2)	road(p)	bin (a)	monkey (u1)
	Test	book	spoon (u1)	paper (a)	balloons (p)	tree (u2)
	Filler	apple	hat (u1)	duck (u2)	apple (t)	blocks (u3)
	Test	bath	foot (u2)	bowl (p)	soap (a)	sheep (u1)
1	Test	park	swing (a)	button (u1)	pushchair (p)	cushions (u2)
	Filler	boat	boat (t)	tomato (u1)	lion (u2)	doll (u3)
	Test	rabbit	carrot (a)	spade (u1)	rocket (p)	windows (u2)
	Test	coat	pushchair (u2)	cushions (p)	swing (u1)	button (a)
	Filler	towel	Cat(u3)	Towel(t)	Boots(u1)	Raisins(u2)
	Test	shoe	sheep (p)	foot (a)	bowl (u1)	soap (u2)
	Test	photo	brush (u2)	flower (p)	camera (a)	house (u1)
	Filler	potty	fridge (u2)	hand (u3)	bee (u1)	potty (t)
	Test	car	monkey (u2)	cup (p)	road (a)	bin (u1)
2	Test	squirrel	tree (a)	spoon (p)	paper (u1)	balloons (u2)
	Filler	sandwich	bed (u1)	dress (u2)	aeroplane (u3)	sandwich (t)
	Test	hair	house (p)	brush (a)	flower (u1)	camera (u2)
	Test	banana	bin (p)	monkey (a)	cup (u1)	road (u2)
	Filler	bag	bag (t)	tiger (u1)	cake (u2)	helicopter (u3)
	Test	farm	soap (u1)	sheep (a)	foot (p)	bowl (u2)
	Test	curtain	windows (a)	carrot (p)	spade (u2)	rocket (u1)
	Filler	fire engine	hoover (u3)	fire engine (t)	teddy (u1)	cucumber (u2)
	Test	table	balloons (u1)	tree (p)	spoon (a)	paper (u2)
3	Test	baby	button (p)	pushchair (a)	cushions (u1)	swing (u2)
	Filler	strawberry	digger (u2)	telephone (u3)	strawberry (t)	penguin (u1)
	Test	party	paper (p)	balloons (a)	tree (u1)	spoon (u2)
	Test	cereal	bowl (a)	soap (p)	sheep (u2)	foot (u1)
	Filler	dog	tractor (u1)	clock (u2)	peas (u3)	dog (t)
	Test	milk	road (u1)	bin (u2)	monkey (p)	cup (a)
	Test	sofa	cushions (a)	swing (p)	button (u2)	pushchair (u1)
	Filler	horse	horse (t)	bib (u1)	pen (u2)	icecream (u3)
	Test	key	camera (p)	house (a)	brush (u1)	flower (u2)
4	Test	watering can	rocket (u2)	windows (p)	carrot (u1)	spade (a)

Note: p = phonological distractor, a = thematic distractor, u = unrelated distractors, t = target

Table B3

Experiment 2–Test List 2.

List 2			Picture Position			
			<i>Upper Left</i>	<i>Upper Right</i>	<i>Lower Left</i>	<i>Lower Right</i>
Block	Type	Target	Pictures (Picture Role)			
0	Filler	grape	highchair (u1)	train (u2)	fish (u3)	grape (t)
	Filler	boat	tomato (u1)	lion (u2)	doll (u3)	boat (t)
	Test	rabbit	spade (u1)	rocket (p)	windows (u2)	carrot (a)
	Test	coat	cushions (p)	swing (u1)	button (a)	pushchair (u2)
	Filler	towel	Towel(t)	Boots(u1)	Raisins(u2)	Cat(u3)
	Test	shoe	foot (a)	bowl (u1)	soap (u2)	sheep (p)
	Test	photo	flower (p)	camera (a)	house (u1)	brush (u2)
	Filler	potty	hand (u3)	bee (u1)	potty (t)	fridge (u2)
	Test	car	cup (p)	road (a)	bin (u1)	monkey (u2)
	Test	squirrel	spoon (p)	paper (u1)	balloons (u2)	tree (a)
2	Filler	sandwich	dress (u2)	aeroplane (u3)	sandwich (t)	bed (u1)
	Test	hair	brush (a)	flower (u1)	camera (u2)	house (p)
	Test	banana	monkey (a)	cup (u1)	road (u2)	bin (p)
	Filler	bag	tiger (u1)	cake (u2)	helicopter (u3)	bag (t)
	Test	farm	sheep (a)	foot (p)	bowl (u2)	soap (u1)
	Test	curtain	carrot (p)	spade (u2)	rocket (u1)	windows (a)
	Filler	fire engine	fire engine (t)	teddy (u1)	cucumber (u2)	hoover (u3)
	Test	table	tree (p)	spoon (a)	paper (u2)	balloons (u1)
	Test	baby	pushchair (a)	cushions (u1)	swing (u2)	button (p)
	Filler	strawberry	telephone (u3)	strawberry (t)	penguin (u1)	digger (u2)
3	Test	party	balloons (a)	tree (u1)	spoon (u2)	paper (p)
	Test	cereal	soap (p)	sheep (u2)	foot (u1)	bowl (a)
	Filler	dog	clock (u2)	peas (u3)	dog (t)	tractor (u1)
	Test	milk	bin (u2)	monkey (p)	cup (a)	road (u1)
	Test	sofa	swing (p)	button (u2)	pushchair (u1)	cushions (a)
	Filler	horse	bib (u1)	pen (u2)	icecream (u3)	horse (t)
	Test	key	house (a)	brush (u1)	flower (u2)	camera (p)
	Test	watering can	windows (p)	carrot (u1)	spade (a)	rocket (u2)
	Filler	sock	sock (t)	tV (u1)	bicycle (u2)	pear (u3)
	Test	sky	rocket (a)	windows (u1)	carrot (u2)	spade (p)
4	Test	butterfly	camera (u1)	house (u2)	brush (p)	flower (a)
	Filler	chicken	bus (u3)	chicken (t)	nappy (u1)	fork (u2)
	Test	rubbish	road(p)	bin (a)	monkey (u1)	cup (u2)
	Test	book	paper (a)	balloons (p)	tree (u2)	spoon (u1)
	Filler	apple	duck (u2)	apple (t)	blocks (u3)	hat (u1)
	Test	bath	bowl (p)	soap (a)	sheep (u1)	foot (u2)
	Test	park	button (u1)	pushchair (p)	cushions (u2)	swing (a)
	Test	sky	rocket (a)	windows (u1)	carrot (u2)	spade (p)
	Test	butterfly	camera (u1)	house (u2)	brush (p)	flower (a)
	Filler	chicken	bus (u3)	chicken (t)	nappy (u1)	fork (u2)

Note: p = phonological distractor, a = thematic distractor, u = unrelated distractors, t = target

Table B4

Experiment 2–Test List 3.

List 3			Picture Position			
			Upper Left	Upper Right	Lower Left	Lower Right
Block	Type	Target	Pictures (Picture Role)			
0	Filler	grape	train (u2)	fish (u3)	grape (t)	highchair (u1)
	Filler	sandwich	aeroplane (u3)	sandwich (t)	bed (u1)	dress (u2)
	Test	hair	flower (u1)	camera (u2)	house (p)	brush (a)
	Test	banana	cup (u1)	road (u2)	bin (p)	monkey (a)
	Filler	bag	cake (u2)	helicopter (u3)	bag (t)	tiger (u1)
	Test	farm	foot (p)	bowl (u2)	soap (u1)	sheep (a)
	Test	curtain	spade (u2)	rocket (u1)	windows (a)	carrot (p)
	Filler	fire engine	teddy (u1)	cucumber (u2)	hoover (u3)	fire engine (t)
	Test	table	spoon (a)	paper (u2)	balloons (u1)	tree (p)
3	Test	baby	cushions (u1)	swing (u2)	button (p)	pushchair (a)
	Filler	strawberry	strawberry (t)	penguin (u1)	digger (u2)	telephone (u3)
	Test	party	tree (u1)	spoon (u2)	paper (p)	balloons (a)
	Test	cereal	sheep (u2)	foot (u1)	bowl (a)	soap (p)
	Filler	dog	peas (u3)	dog (t)	tractor (u1)	clock (u2)
	Test	milk	monkey (p)	cup (a)	road (u1)	bin (u2)
	Test	sofa	button (u2)	pushchair (u1)	cushions (a)	swing (p)
	Filler	horse	pen (u2)	icecream (u3)	horse (t)	bib (u1)
	Test	key	brush (u1)	flower (u2)	camera (p)	house (a)
4	Test	watering can	carrot (u1)	spade (a)	rocket (u2)	windows (p)
	Filler	sock	tV (u1)	bicycle (u2)	pear (u3)	sock (t)
	Test	sky	windows (u1)	carrot (u2)	spade (p)	rocket (a)
	Test	butterfly	house (u2)	brush (p)	flower (a)	camera (u1)
	Filler	chicken	chicken (t)	nappy (u1)	fork (u2)	bus (u3)
	Test	rubbish	bin (a)	monkey (u1)	cup (u2)	road(p)
	Test	book	balloons (p)	tree (u2)	spoon (u1)	paper (a)
	Filler	apple	apple (t)	blocks (u3)	hat (u1)	duck (u2)
	Test	bath	soap (a)	sheep (u1)	foot (u2)	bowl (p)
1	Test	park	pushchair (p)	cushions (u2)	swing (a)	button (u1)
	Filler	boat	lion (u2)	doll (u3)	boat (t)	tomato (u1)
	Test	rabbit	rocket (p)	windows (u2)	carrot (a)	spade (u1)
	Test	coat	swing (u1)	button (a)	pushchair (u2)	cushions (p)
	Filler	towel	Boots(u1)	Raisins(u2)	Cat(u3)	Towel(t)
	Test	shoe	bowl (u1)	soap (u2)	sheep (p)	foot (a)
	Test	photo	camera (a)	house (u1)	brush (u2)	flower (p)
	Filler	potty	bee (u1)	potty (t)	fridge (u2)	hand (u3)
	Test	car	road (a)	bin (u1)	monkey (u2)	cup (p)
2	Test	squirrel	paper (u1)	balloons (u2)	tree (a)	spoon (p)

Note: p = phonological distractor, a = thematic distractor, u = unrelated distractors, t = target

Table B5

Experiment 2–Test List 4.

List 4			Picture Position			
			Upper Left	Upper Right	Lower Left	Lower Right
Block	Type	Target	Pictures (Picture Role)			
0	Filler	grape	fish (u3)	grape (t)	highchair (u1)	train (u2)
	Filler	strawberry	penguin (u1)	digger (u2)	telephone (u3)	strawberry (t)
	Test	party	spoon (u2)	paper (p)	balloons (a)	tree (u1)
	Test	cereal	foot (u1)	bowl (a)	soap (p)	sheep (u2)
	Filler	dog	dog (t)	tractor (u1)	clock (u2)	peas (u3)
	Test	milk	cup (a)	road (u1)	bin (u2)	monkey (p)
	Test	sofa	pushchair (u1)	cushions (a)	swing (p)	button (u2)
	Filler	horse	icecream (u3)	horse (t)	bib (u1)	pen (u2)
	Test	key	flower (u2)	camera (p)	house (a)	brush (u1)
1	Test	watering can	spade (a)	rocket (u2)	windows (p)	carrot (u1)
	Filler	sock	bicycle (u2)	pear (u3)	sock (t)	tV (u1)
	Test	sky	carrot (u2)	spade (p)	rocket (a)	windows (u1)
	Test	butterfly	brush (p)	flower (a)	camera (u1)	house (u2)
	Filler	chicken	nappy (u1)	fork (u2)	bus (u3)	chicken (t)
	Test	rubbish	monkey (u1)	cup (u2)	road(p)	bin (a)
	Test	book	tree (u2)	spoon (u1)	paper (a)	balloons (p)
	Filler	apple	blocks (u3)	hat (u1)	duck (u2)	apple (t)
	Test	bath	sheep (u1)	foot (u2)	bowl (p)	soap (a)
2	Test	park	cushions (u2)	swing (a)	button (u1)	pushchair (p)
	Filler	boat	doll (u3)	boat (t)	tomato (u1)	lion (u2)
	Test	rabbit	windows (u2)	carrot (a)	spade (u1)	rocket (p)
	Test	coat	button (a)	pushchair (u2)	cushions (p)	swing (u1)
	Filler	towel	Raisins(u2)	Cat(u3)	Towel(t)	Boots(u1)
	Test	shoe	soap (u2)	sheep (p)	foot (a)	bowl (u1)
	Test	photo	house (u1)	brush (u2)	flower (p)	camera (a)
	Filler	potty	potty (t)	fridge (u2)	hand (u3)	bee (u1)
	Test	car	bin (u1)	monkey (u2)	cup (p)	road (a)
3	Test	squirrel	balloons (u2)	tree (a)	spoon (p)	paper (u1)
	Filler	sandwich	sandwich (t)	bed (u1)	dress (u2)	aeroplane (u3)
	Test	hair	camera (u2)	house (p)	brush (a)	flower (u1)
	Test	banana	road (u2)	bin (p)	monkey (a)	cup (u1)
	Filler	bag	helicopter (u3)	bag (t)	tiger (u1)	cake (u2)
	Test	farm	bowl (u2)	soap (u1)	sheep (a)	foot (p)
	Test	curtain	rocket (u1)	windows (a)	carrot (p)	spade (u2)
	Filler	fire engine	cucumber (u2)	hoover (u3)	fire engine (t)	teddy (u1)
	Test	table	paper (u2)	balloons (u1)	tree (p)	spoon (a)
	Test	baby	swing (u2)	button (p)	pushchair (a)	cushions (u1)

Note: p = phonological distractor, a = thematic distractor, u = unrelated distractors, t = target

Appendix C

Level-1 model:

$$Y_{ij} = \alpha_{0i} + \beta_{1i}Time_{ij} + \beta_{2i}Time_{ij}^2 + \beta_{3i}Time_{ij}^3 + \varepsilon_{ij} \quad (1)$$

Level-2 models:**Intercept:**

$$\alpha_{0i} = \gamma_{00} + \gamma_{01}OBJECT + \gamma_{02}AGE/VOCAB + \gamma_{03}(OBJECT * AGE/VOCAB_i) + \zeta_{0i} \quad (2)$$

Linear:

$$\beta_{1i} = \gamma_{10} + \gamma_{11}OBJECT + \gamma_{12}AGE/VOCAB + \gamma_{13}(OBJECT * AGE/VOCAB_i) + \zeta_{1i} \quad (3)$$

Quadratic:

$$\beta_{2i} = \gamma_{20} + \gamma_{21}OBJECT + \zeta_{2i} \quad (4)$$

Cubic:

$$\beta_{3i} = \gamma_{30} + \gamma_{31}OBJECT + \zeta_{3i} \quad (5)$$

AGE/VOCAB = Age or Receptive Vocabulary or Expressive Vocabulary

The first number index indicates the polynomial order (0 = intercept, 1 = linear, 2 = quadratic, 3 = cubic).

The second number index indicates the predictor (1 = Object, 2 = D, 3 = Object*Age/Vocab).

Subscript *i* indicates an individual participant.

Subscript *j* indicates an particular measurement occasion.

Appendix D

Figure D1. Pilot Experiment. Growth curve analysis of (a) 24-month-old and (b) 30-month-old participants. Data shown in points, the model shown in lines.

Table D1.

Pilot Experiment. Growth Curve Analysis.

Fixed Effects	Model Terms	Estimates	SE	z	p
Target–Unrelated	Intercept	1.19	0.07	18.38	< .001***
	Linear	2.45	0.23	10.81	< .001***
	Quadratic	-0.26	0.17	-1.56	.117
	Cubic	-0.27	0.16	-1.65	.100
Phonological–Unrelated	Intercept	0.04	0.08	0.48	.619
	Linear	-0.78	0.26	-3.04	.002**
	Quadratic	0.10	0.19	0.52	.605
	Cubic	0.24	0.18	1.31	.192
Taxonomic–Unrelated	Intercept	0.05	0.07	0.70	.484
	Linear	0.70	0.25	2.76	.006**
	Quadratic	-0.03	0.18	-0.18	.856
	Cubic	-0.10	0.18	-0.55	.585
24-month-olds–30-month-olds	Intercept	0.31	0.12	2.68	.007**
	Linear	0.08	0.26	0.30	.7642
Target–Unrelated *24-months – Target–Unrelated *30-months	Intercept	-0.66	0.10	-6.83	< .001***
	Linear	-0.10	0.33	-0.31	.756
Phonological–Unrelated *24-months – Phonological–Unrelated *30-months	Intercept	-0.22	0.11	-2.00	.044*
	Linear	0.62	0.37	1.65	.098
Taxonomic–Unrelated *24-months – Taxonomic–Unrelated *30-months	Intercept	-0.01	0.10	-0.11	.912
	Linear	0.11	0.36	0.30	.762

*** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$

Appendix E

We collected visual similarity ratings based on the methods in De Groot et al. (2015) using an online survey. Participants were instructed to rate the visual similarity of a written word (the target) and a picture (the phonological, semantic or unrelated distractor) purely based on visual similarity using an 11-point Likert scale (0 = No similarity at all; 10 = Very Similar). Instructions were “On each page, please rate the visual similarity of the two items depicted by the word and image: 0 means no similarity at all, 10 means very similar. Please ignore the other dimensions such as phonological or semantic relations of the items”. In total, 16 adults were recruited from the University of Oxford. Eight were native British English speakers, and the rest self-rated as fluent English speakers who have lived in the UK for an average of 2.7 years ($SD = 1$, Range = 1-4 years). The following tables present the results of the similarity rating studies. Comparisons of visual similarity ratings were carried out using the nonparametric Wilcoxon signed-rank test.

Table E1. Experiment 1 Visual Similarity Ratings

<i>Phonological vs. Taxonomic: Visual Similarity Ratings (N=16)</i>			
	Phonological	Taxonomic	Unrelated
Mean Rating (standard deviation)	2.06 (1.43)	2.68 (1.30)	2.05 (1.56)
Comparison with Taxonomic	$Z = -2.85, p = .004^*$	--	--
Comparison with Unrelated	$Z = -.568, p = .570$	$Z = -2.74, p = .006^{**}$	--
*** $p < .001$, ** $p < .01$, * $p < .05$			

Table E2. Experiment 2 Visual Similarity Ratings

<i>Phonological vs. Thematic: Visual Similarity Ratings (N = 16)</i>			
	Phonological	Thematic	Unrelated
Mean Rating (standard deviation)	2.08 (1.13)	3.92 (2.16)	2.11 (1.23)
Comparison with Thematic	$Z = -3.26, p = .001^{***}$	--	--
Comparison with Unrelated	$Z = -.03, p = .979$	$Z = -3.15, p = .002^{**}$	--
*** $p < .001$, ** $p < .01$, * $p < .05$			

As seen in Tables E1 and E2, the visual similarity ratings for between the target and the phonological or unrelated distractors did not differ significantly in both experiments. We can, therefore, be relatively confident that toddlers' preference for the phonological distractor over the unrelated

distractor was not driven by visual similarity with the target. However, the visual similarity ratings for both taxonomic and thematic distractors were significantly higher than that for the phonological and unrelated distractors. To some extent, we expected the taxonomically-related pictures to be visually more similar to the target, since taxonomic category members are defined by shared features. However, we did not expect the thematic distractors to be visually similar to the target. For example, the adult participants rated Sky-Rocket (thematically-related, rating: 3.19) to be visually more similar than Sky-Spade (phonologically-related, rating: 1.81) and Sky-Carrot (unrelated, rating: 1.69). Note that rocket, spade and carrot are visually-similar, all three are long objects. It is, therefore, peculiar that adult participants would find the rocket to be visually more similar to the sky.

Interestingly, De Groot et al. (2015) observed the same pattern of results in their adult visual similarity rating study: visually-related pictures are rated as semantically more similar, and conversely, semantically-related pictures were rated as visually more similar. De Groot et al. (2015) suggested several reasons, including the inherent difficulty for one to completely separate semantic and visual representations, and in the context of a rating experiment, it may be difficult for participants to follow the task instructions and respond purely based on one dimension.